BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions

The robustness of conditional logit for binary response panel data models with serial correlation

Do Won Kwak, Korea University
Robert S. Martin, U.S. Bureau of Labor Statistics
Jeffrey M. Wooldridge, Michigan State University

Working Paper 502
August 2018

The robustness of conditional logit for binary response panel data models with serial correlation*

Do Won Kwak† Robert S. Martin‡ Jeffrey M. Wooldridge§

May 30, 2018

Abstract

This paper examines the conditional logit estimator for binary panel data models with unobserved heterogeneity. A key assumption used to derive the conditional logit estimator is conditional serial independence (CI), which is problematic when the underlying innovations are serially correlated. A Monte Carlo experiment suggests that the conditional logit estimator is not robust to violation of the CI assumption. We find that higher persistence and smaller time dimension both increase the magnitude of the bias in slope parameter estimates. We also compare conditional logit to unconditional logit and pooled correlated random effects logit.

Keywords: Panel data; Binary dependent variable; Conditional logit model; Unobserved heterogeneity

JEL Codes: C15, C23, C25

*We are grateful to David Drukker, Todd Elder, Iraj Rahmani, Peter J. Schmidt, and Timothy J. Vogelsang for their helpful suggestions and comments. The views expressed herein are those of the authors and not necessarily those of the Bureau of Labor Statistics or the U.S. Department of Labor.

†Asiatic Research Institute, Korea University, Seoul, 02841, Korea. Tel: +82 2-3290-5259. Fax: +82 2-923-4661 Email: dwkwak@korea.ac.kr

‡Division of Price and Index Number Research, Bureau of Labor Statistics, 2 Massachusetts Ave, NE, Washington, DC 20212, USA. Tel: 202-691-6952. Fax: 202-691-6583. Email: Martin.Robert@bls.gov

§Department of Economics, Michigan State University, East Lansing, MI 48824-1038, USA. Tel: 517-353-5972. Fax: 517-432-1068 Email: wooldri1@msu.edu

1 Introduction

In the context of binary response panel data models, standard maximum likelihood estimators that treat individual heterogeneity as parameters to estimate are known to be inconsistent when the time dimension T is fixed, and the cross section dimension N gets large. The problem arises because the number of unit-specific unobserved effects grows, while information available for estimating the individual-specific parameters does not accumulate.[1] One identification strategy is to assume a logistic distribution of the latent model errors and use the conditional logit approach of Chamberlain (1980).[2] Conditional logit delivers $\sqrt{N}$-consistent estimation of the latent model slope parameters as long as the covariates are strictly exogenous and the outcomes are serially independent conditional on the covariates and heterogeneity. One positive aspect to the method is that the distribution of the unobserved effect is left unspecified. This means we can consistently estimate the partial effects of continuous covariates relative to each other, and we can estimate the effects of the covariates on the log-odds ratio. Unfortunately, the fixed-T consistency result for model coefficients does not extend to average partial effects, as we lack a consistent estimators of the unit-specific unobserved heterogeneity terms.

A key assumption in deriving the conditional MLE in the unobserved effects logit model is conditional independence (CI), which assumes that, given the observed covariates and unobserved heterogeneity, the binary responses are independent over time. This assumption is difficult to justify in many microeconometric applications, as it is tantamount to assuming that time-varying, unobserved innovations are serially independent. For example, why should innovations that affect variables such as labor force participation or welfare participation be independent over time? (Chay and Hyslop, 2014). Other examples include the incidence of external debt crises in developing countries (Hajivassiliou and McFadden, 1998) and habitual smoking behavior (Collado and Browning, 2007). In none of these cases would economic theory imply independence of the underlying innovations. In fact, empirical researchers tend to use clustered standard errors

[1]This situation is called the incidental parameters problem (IPP), described first by Neyman and Scott (1948). See Lancaster (2000) and Arellano and Hahn (2007) for a review.

[2]These types of conditional methods are not available in general nonlinear models and cannot be extended to APE estimation since they do not place any distributional assumption on the unobserved heterogeneity. For general nonlinear models, a more recent literature focuses rather on approximately unbiased estimator than on the estimators with no bias at all. Arellano and Hahn (2007), Hahn and Newey (2004), Arellano (2003), Fernandez-Val (2009) and others provide bias correction methods for nonlinear fixed effects panel data models. An extensive review is provided in Arellano and Hahn (2007).

with linear panel models to account for serial correlation in inference. In the nonlinear case, imposing the CI restriction on the data may be unwarranted and might result in systematic biases in both parameter estimates and APEs.

Naturally, just because the CI assumption is used in deriving the logit conditional MLE – often called the conditional logit (CL) estimator – does not mean the CL estimator is especially sensitive to violations of this assumption. As shown in Wooldridge (1999), the conditional MLE in the Poisson unobserved effects models, which was originally derived under the CI assumption, is consistent in the presence of unrestricted serial dependence in the outcomes in the fixed T, $N \to \infty$ environment. (The conditional MLE is also robust to arbitrary violations of the Poisson assumption.) Unfortunately, no such result has been shown for the conditional logit estimator. As the theoretical literature currently stands, the CI assumption appears necessary for fixed T, $N \to \infty$ consistency of the CL estimator. From the perspective of an applied econometrician, one would want to know the magnitudes of the finite-sample bias when CI fails.

In this paper we fill a gap in the literature and study the robustness of the conditional logit estimator when the CI assumption fails. Generally, it is difficult in nonlinear contexts to show that certain assumptions are necessary for an estimator to be consistent, and it is even harder to characterize the magnitude of the inconsistency using analytical calculations. Instead, we use a Monte Carlo experiment to study the bias in the CL estimator when T is fairly small, as still often happens in micro panel data sets. Specifically, we use data generating processes (DGPs) that satisfy all of the conditions for the conditional logit estimator except CI. The violation of CI is implemented by generating serially correlated errors in the latent variable formulation of the model. We use a Gaussian copula to impose a marginal logistic distribution for the innovations in each period while exploiting a first-order autoregressive structure in the latent errors. We then examine the performance of the conditional logit estimator.

While our primary aim is to study the properties of one of the most widely used estimators for binary panel data models, it makes sense to include a couple of competitors so that we can see if the CL estimator is unusually deficient. In choosing competitors to CL, we want realism in plausible alternative estimation approaches without assuming that the underlying true model is known. We consider two competitors. The first is sometimes called "unconditional logit," which is pooled MLE treating the unobserved heterogeneity as parameters to be estimated.[3] The second competitor is a correlated random effects (CRE) logit model, where we apply the Mundlak (1978)-Chamberlain (1980) device in modeling the distribution of the unobserved effects conditional on the history of covariates. Rather than use the joint MLE under the assumption of conditional independence, we use a simple pooled MLE procedure. There are a few reasons for this. First, as discussed in Wooldridge (2010, Chapter 15), pooled CRE approaches generally have the advantage that they identify the magnitudes of the effects – the so-called average partial effects, or APEs – without the CI assumption. In fact, in the probit case, Wooldridge (2010) shows how scaled parameters are identified by the pooled MLE and these same scaled coefficients identify the average partial effects. Unfortunately, there are no such results for the logit case, regardless of the conditional distribution assumed about the heterogeneity. The reason is that, even if we assume the heterogeneity has a conditional logistic distribution, the composite error term does not have a logistic distribution. Rather than study the CRE probit estimator, which would seem an odd choice when we are primarily interested in properties of the conditional logit estimator, we use a CRE approach where the response probability is necessarily misspecified. In other words, the functional form of the true response probability is unknown, but we assume it is logit.

Our inclusion of the CRE logit approach reflects the idea that in modern applications of binary response panel data models, researchers are usually interested in obtaining magnitudes of the covariate effects on the response probabilities – that is, average partial effects (APEs). Strictly speaking, conditional logit does not yield APE estimates because we do not have a distribution of the unobserved heterogeneity to integrate out of the response probability. Nevertheless, because it is possible that the slope coefficients on the covariates can be estimated with conditional logit, we can define estimated APEs by inserting estimates of the unobserved effects. To explore this possibility, we examine a hybrid approach whereby the heterogeneity is estimated by unconditional MLE when the slope parameters are restricted to be equal to their conditional logit estimates. The conditional logit slopes and second stage estimated heterogeneity are then

[3] We are hardly the first to simulate the performance of unconditional logit (see Greene (2004), for example), though we are unaware of any published studies that relax CI. For theoretical properties, see Arellano and Hahn (2007) and Andersen (1973). For the case of $T = 2$, Abrevaya (1997) shows the bias of unconditional MLE is 100%. This estimator is computationally intenstive, too.

combined to form APE estimates. Corrections for the incidental parameters problem in these APE estimates are available from Hahn and Newey (2004) and Fernandez-Val (2009), and they also assume CI. We compare this hybrid approach to the APE from unconditional logit, fixed effects estimation of the linear probability model, and CRE logit. In the CRE logit case, it will not make sense to compare magnitudes of the coefficients because they have been implicitly scaled.

Our main simulation findings can be summarized here. We find the conditional logit estimator has large biases for the index slope when the CI assumption is violated, though the bias is still less than for unconditional logit. The bias is greater for panels with either stronger serial correlation or a smaller time dimension. The size distortion of inference (the bias in the rejection frequency) with conditional logit is very severe for the data with moderate-to-severe serial correlation, and it increases with the size of cross-section. We also find that the APEs estimated with the hybrid conditional logit approach are significantly biased when T is small, even when correcting for incidental parameters. When T is large, the biases are fairly small no matter the amount of serial correlation. Still, the other APE estimators – those for unconditional logit and CRE logit – tend to have better finite sample properties for our particular data generating process.

The findings in this paper have important consequences for applied researchers who must choose among estimates of different approaches to binary response panel data models. The conditional logit approach is often hailed as being superior to the correlated random effects approach (CRE) because the former leaves the heterogeneity distributions – both unconditional and conditional – unrestricted, while CRE approaches restrict the conditional distribution in some way, often fully parametric. What is often overlooked as that the CL approach imposes an important restriction along a different direction, namely, the CI assumption. CRE approaches relax this assumption entirely when the focus is on APEs. Whether one should use logit, probit, or more flexible models in the CRE approach is an area that deserves further research. Approaches that treat heterogeneity as parameters to estimate and then seek to remove the resulting bias are promising, but they can only be shown to have good properties when T is sufficiently large.

The rest of paper is organized as follows. In Section 2, we explicitly state the model and ideal assumptions for conditional logit model. Section 3 describes the data generating procedure and how we use a copula to relax CI. Section 4 shows the finite sample results of conditional logit and competing methods. Section 5 illustrates these methods using data on womens welfare participation, while Section 6 concludes.

2 Unobserved Effects Logit Model

We are interested in a binary response variable, y_{it}, that is generated by an underlying latent variable model.

Assumption 2.1

$$y_{it} = 1\left[\boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it} > 0\right],\ i = 1, 2, \ldots, N;,\ t = 1, 2, \ldots, T,$$

where $1[\ ]$ is the indicator function, $\mathbf{x}_{it}$ is a $1 \times K$ vector of explanatory variables, c_i is unobserved and thought to be correlated with $\boldsymbol{x}_{it}$, and u_{it} is the idiosyncratic error. We consider the setting where N is large and T, while being at least two, is fixed in the asymptotic anlaysis.

Assumption 2.2 *The idiosyncratic error u_{it} is marginally distributed standard logistic with CDF*

$$\Lambda(u) = \frac{\exp(u)}{1 + \exp(u)}.$$

Under (2.1) and (2.2), the marginal distribution of y_{it} conditional on $\mathbf{x}_{it}$ and c_i, is logit, with probability of success given by

$$P(y_{it} = 1|\mathbf{x}_{it}, c_i) = \Lambda(\mathbf{x}_{it}\beta + c_i). \tag{1}$$

The presence of c_i that can be correlated with $\mathbf{x}_{it}$ necessitates further assumptions order to identify and estimate $\boldsymbol{\beta}$. In this paper we impose the common assumption of strict exogeneity of $\{\mathbf{x}_{it} : t = 1, 2, ..., T\}$ conditional on c_i.

Assumption 2.3 (Strict Exog. Conditional on the Unobserved Heterogeneity)

$$P(y_{it} = 1|\mathbf{x}_i, c_i) = P(y_{it} = 1|\mathbf{x}_{it}, c_i),\ t = 1, \ldots, T$$

where $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, ..., \mathbf{x}_{iT})$.

In terms of the idiosyncratic error, (2.3) means that u_{it} is independent of the entire history of covariates, $\{\mathbf{x}_{it} : t = 1, 2, ..., T\}$.

Assumption 2.4 (Conditional Independence)

$$D(y_{i1,}y_{i2}, ..., y_{iT}|\mathbf{x}_i, c_i) = \prod_{t=1}^{T} D(y_{it}|\mathbf{x}_{it}, c_i)$$

The notation $D(\cdot|\cdot)$ denotes conditional distribution.

Under these assumptions, along with weak regularity conditions discussed in Chamberlain (1980) and Andersen (1970), conditional logit allows us to estimate $\boldsymbol{\beta}$. Conditioning on $n_i \equiv \sum_{t=1}^{T} y_{it}$ drops c_i from the likelihood, separating it from the estimation of the structural parameters (Andersen (1973), Chamberlain (1984) and Wooldridge (2010)). The strength of the approach is we need not make any substantive assumptions about $D(c_i|\mathbf{x}_i)$. As far as we know from a theoretical perspective, the generality of allowing any relationship between c_i and $\mathbf{x}_i$ comes at the the expense of imposing the the conditional independence assumption. The main purpose of this paper is to determine the practical importance of (2.4), which requires that $\{u_{it}\}_{t=1}^{T}$ is serially independent, in estimating both parameters and average partial effects.

The unconditional logit approach treats the c_i as parameters to estimate along with β, and maximizes the pooled log likelihood across i and t. The CRE approach we use imposes the Mundlak-Chamberlain device

$$c_i = \psi + \bar{\mathbf{x}}_i \xi + a_i$$

and then uses the pooled logit likelihood with response probability

$$P(y_{it} = 1|\mathbf{x}_i) = \Lambda(\eta_a + \mathbf{x}_{it}\beta_a + \bar{\mathbf{x}}_i \xi_a)$$

where the a subscript denotes that the parameters are scaled relative to the original parameteres. In the context of CRE logit, we cannot derive the precise scaling because the true response probability does not even have a logit form. Thus, we are intentionally using a misspecified response probability, but we are relaxing the CI assumption. The relationship between the heterogeneity and the covariates is captured through the Mundlak-Chamberlain device.

2.1 Average Partial Effects

In our simulation study we consider both parameters and an average partial effect. If x_j is continuous, its APE, averaged across all time periods, is

$$APE_j \equiv \beta_j E\left[\frac{1}{T}\sum_{t=1}^{T}\lambda(\mathbf{x}_{it}\beta + c_i)\right], \tag{2}$$

where

$$\lambda(u) \equiv \frac{\mathrm{d}\Lambda}{\mathrm{d}u}(u)$$

and the expected value is with respect to the joint distribution of $(\mathbf{x}_i, c_i)$. The estimate of the APE differs across the three methods. If $\hat{\beta}_{CL}$ are the conditional logit estimates, then we plug this into the log-likelihood for each i, and obtain estimates $\hat{c}_{i,CL}$ for each i. The estimated APE is

$$\widehat{APE}_{j,CL} \equiv \hat{\beta}_{j,CL}\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda(\mathbf{x}_{it}\hat{\beta}_{CL} + \hat{c}_{i,CL})\right] \tag{3}$$

In the unconditional logit case we jointly obtain $\hat{\beta}_{UL}$ and $\hat{c}_{i,UL}$ and insert these into (3).[4] In the CRE logit case, we obtain $\hat{\eta}_{a,CRE}$, $\hat{\beta}_{a,CRE}$, and $\hat{\xi}_{a,CRE}$ and then compute

$$\widehat{APE}_{j,CRE} \equiv \hat{\beta}_{aj,CRE}\left[(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\lambda(\hat{\eta}_{a,CRE} + \mathbf{x}_{it}\hat{\beta}_{a,CRE} + \bar{\mathbf{x}}_i\hat{\xi}_{a,CRE})\right]$$

The inclusion of $\bar{\mathbf{x}}_i$ serves as a proxy to c_i, as discussed in Wooldridge (2010).

Even when β is estimated consistently, we expect the IPP to bias $\widehat{APE}_{j,CL}$ due to the inclusion of the estimated incidental parameters $\hat{c}_{i,CL}$. Therefore, we employ Fernandez-Val's (2009) (FV) bias correction for the APE.[5] The corrected APE has been shown to be approximately unbiased when $\rho = 0$ and T is moderately large, but less is known about its properties when $\rho \neq 0$.

Table 1 summarizes the features of the estimators we consider in this paper. Ex ante, no

[4] The relevant sample size for this and the unconditional logit APE is N, rather than the number of cross-section units with nonconstant y_t. The estimated heterogeneity of an individual j with constant y_{jt} is unbounded, meaning its contribution to the sample analog of (2) identically zero (Alexander and Breunig, 2014).

[5] Hahn and Newey (2004) also proposed similar corrections.

procedure strictly dominates the others, so it is up to the analyst to judge the relative importance of the different factors. Our experiment, described in the next section, aims to shed light on the role of conditional independence.

Table 1: Considered estimators for unobserved effects logit model

	$P(y_{it}=1\|\cdot)$ correct?	$P(y_{it}=1\|\cdot)$ in $(0,1)$?	Restricts $D(c_i\|\boldsymbol{x}_i)$?	Assumes cond. indep.?	IPP?
Conditional logit	Yes	Yes	No	Yes	No
Unconditional logit	Yes	Yes	No	No	Yes
CRE logit, pooled MLE	No	Yes	Yes	No	No
FE OLS (LPM)	No	No	No	No	No

3 Description of the Monte Carlo Experiment

This section introduces the details of data generating processes in simulation for studying the robustness of conditional logit to violation of the conditional independence assumption. The key is to generate variables that satisfy all of the ideal conditions for conditional logit except CI. We generate data based on a simple model with one continuous covariate that is correlated with the heterogeneity. We then test the performance of conditional logit and the competing estimators described in the previous section. All the results presented in this paper are based on 2,000 replications.

3.1 Model Design

We generate data $\{y_{it}, x_{it}, c_i, u_{it}\}$ for $t = 1, 2, .., T$ and $i = 1, 2, ..., N$ based on model (2.1). We follow Heckman (1981), Greene (2004) and Fernandez-Val (2009) for the design of the DGPs in this section. The latent equation contains one continuous regressor as well as time-constant unobserved heterogeneity.[6] In other designs with binary regressors, we found the results to be qualitatively similar. We draw the four items with 2,000 replications.

- $\{x_{it}\}_{t=1}^{T}$ are independent, identically distributed as $Normal(0, 1)$.

[6] Devroye (1986) and Johnson (1987) contain extensive methods of generating multivariate random variables and vectors from non-normal distributions.

- Unobserved heterogeneity:

$$c_i = \frac{x_{i1} + x_{i2} + ... + x_{ik}}{\sqrt{2k}} + \frac{a_i}{\sqrt{2}} \quad \text{where } a_i \sim \text{Logistic}(0, \frac{\pi^2}{3}) \text{ and } k = \frac{T}{2}.$$

- We generate each u_{it} as described in Section 3.2, such that it has standard logistic marginal distribution, but $Cov(u_{it}, u_{is}) = \rho^{|t-s|}$ for $t \neq s$.

- We then generate y_{it} as follows:

$$y_{it} = 1[\beta x_{it} + c_i + u_{it} \geq 0],$$

where we set $\beta = 1$.

3.2 Generation of Correlated Logistic Errors

We introduce serial correlation among the marginal logistic errors $\boldsymbol{u}_i$ using a Gaussian copula. In general, specifying variables from multivariate distributions can be very complicated as we have to fully define the joint distribution. However, using a copula, we can need only specify the marginal logistic distribution of each variable, as the copula takes care of the generating serial dependence.[7] In the following, we suppress the i subscripts.

Let $\mathbf{z} = (z_1, z_2, ..., z_T)'$ be drawn drawn from T-variate multivariate normal distribution with an AR(1) correlation structure:

$$Cov(\mathbf{z}) = \begin{pmatrix} 1 & \rho & \cdots & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \rho^2 \\ \rho^{T-2} & \rho^{T-3} & \cdots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \cdots & \cdots & 1 \end{pmatrix} \tag{4}$$

We obtain $(u_1, u_2, ..., u_T)$ by the inverse probability transform of the marginal distributions:

$$u = \Lambda^{-1}(\Phi(z_t)), \tag{5}$$

[7]See Greene (2012), 12.2.2 for a more general discussion of modeling with copula functions.

where Λ is logistic CDF and Φ is standard normal CDF. Specifically, $\mathbf{u} = (u_1, u_2, ..., u_T)$ is generated from the following algorithm:

(Step 1) Draw $z_1, z_2, ..., z_T$ from T-variate normal distribution with the AR(1) dependence structure in (4).

(Step 2) Generate correlated Uniform(0,1) random variables as $e_t = \Phi(z_t)$.

(Step 3) Generate the u_t as haveing the standard logistic transformation as

$$u_t = \Lambda^{-1}(e_t) = \Lambda^{-1}(\Phi(z_t))$$

where

$$\Lambda^{-1}(e) = \log\left(\frac{e}{1-e}\right)$$

We repeat steps of (1), (2), and (3) for random draws $i = 1, 2, ..., N$.

4 Results

We study the performance of conditional logit and the competing estimators as T, N, and ρ change. We are particularly interested in the typical micro panel setting of small T, large N, and high serial correlation. Tables 5 through 28 present detailed results. For either β or the APE, we report the empirical mean of the point estimate, the empirical standard deviation of the point estimate, the ratio of the mean point estimate of the standard error to the standard deviation (SE/SD), and the rejection frequency $[1(p_{val} < 0.05)]$ of a true null hypothesis. Standard errors are clustered by cross-sectional unit to account for serial correlation. We also normalize APE estimates to have a true value of 1 using a single draw with $N = 1,000,000$ to estimate the true mean.

A common thread underlying the finite sample results for conditional and unconditional logit concerns the estimation sample. Both estimators only use cross sectional units for which the outcome variable varies over time. This means that even though we experiment with nominal sample sizes of $N \in \{100, 200, 400, 600\}$, the effective sample size can be much smaller. Table 2 gives the proportion of observations actually used in maximum likelihood estimation, that is,

Table 2: MLE and CMLE Effective Cross Section Size (proportion)

	$\rho = 0$	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
$T = 2$	0.3631	0.3317	0.2984	0.2621	0.2212
$T = 3$	0.5582	0.5227	0.4785	0.4248	0.3575
$T = 4$	0.6664	0.6346	0.5908	0.5314	0.4491
$T = 5$	0.7413	0.7137	0.6733	0.6134	0.5227
$T = 6$	0.7898	0.7666	0.7308	0.6734	0.5794
$T = 7$	0.8282	0.8081	0.7758	0.7222	0.6276
$T = 8$	0.8537	0.8364	0.8083	0.7594	0.6671

Empirical mean probability that a cross sectional unit's response varies over time. Estimated with N=600, 2,000 replications.

excluding observations for which the outcomes are all 1 or all 0.[8] For instance, in the $T = 2$ case, even when CI holds, the presence of heterogeneity (which causes y_{it} to be unconditionally serially correlated) reduces the proportion of included observations from 50% (which we would expect if y_{it} were i.i.d.) to 36%. As one would expect, increasing conditional serial correlation through ρ also increases the likelihood an observation's response will be constant over time and therefore reduces the effective sample size. As we discuss in the following subsections, we expect that this mechanism exacerbates the effect of higher ρ on the finite sample properties of conditional and unconditional logit, particularly when T is small.

4.1 Comparison of Slope Estimators

4.1.1 Conditional Logit

As predicted by theory, when CI holds ($\rho = 0$), the conditional logit estimator shows little finite sample bias for the coefficient β as long as the sample size is not too small. For example, when $N = 200$ (as shown in panel (a) of Figure 1), the empirical bias is about 5% when $T = 2$, but falls to less than 2% for $T \geq 3$. For $N \geq 400$, the bias is less than 1% for all $T \geq 3$. However, substantial bias is evident when ρ is non-zero, and we highlight three features using results from the $N = 200$ case. First, the magnitude of the bias is greater with higher values of ρ. For instance, when $T = 5$, the bias is 6.6% when serial correlation is mild ($\rho = 0.2$), but jumps to 34.6% when serial correlation is moderately strong ($\rho = 0.6$). Second, bias is greater the smaller is T. In the $T = 2$ case, the bias is 19.4% when $\rho = 0.2$, and 79.3% when $\rho = 0.6$, which is

[8] Note CRE and FE OLS use all observations including those that do not change response over time.

substantially more severe than the $T = 5$ estimates. Finally, when $T > 2$, the bias changes little when we increase the cross section size N (as we would expect if N is large enough for the Monte Carlo average to be close to the probability limit). To illustrate, in the $T = 5$, $\rho = 0.6$ case, the bias is 34.6% when $N = 100$ and only slightly lower (33.8%) when $N = 600$. We find increasing N reduces bias more when $T = 2$, particularly when serial correlation is strong. But while bias is reduced by about 200 percentage points in the $\rho = 0.8$ case when we increase N from 100 to 200, it is only reduced by about seven percentage points when we increase N from 400 to 600.

As one would expect, the empirical standard deviations fall as either T or N increase. Disproportionate increases in sampling error can occur with increases in ρ when T is small. For example when $N = 200$ and $T = 2$, the standard deviation increases from 0.29 when $\rho = 0$ to 4.72 when $\rho = 0.8$ even though the mean estimate of β only triples. When $T = 8$, however the increase in standard deviation is only from 0.08 to 0.12. At least some of the increase in sampling variation is likely due to decreases in the effective sample size. Somewhat surprisingly, the clustered standard errors appear to do a decent job accounting for the serial correlation as long as $T \geq 3$, despite depending on biased estimates of β. The ratio of the mean standard error to the empirical standard deviation is within 5% of unity in most cases. Problems with inference, therefore are driven primarily by the bias as opposed to the standard errors. As expected, we find no size distortion when $\rho = 0$ in that, for all T, the empirical rejection probability (of the hypothesis test of $H_0 : \beta = 1$) is statistically indistinguishable from 0.05. Not surprisingly, however, the experiment implies substantial over-rejection when $\rho > 0$. Moreover, the size distortion increases with N (as standard errors become smaller and the t statistic is centered about the incorrect value), and can be significant even with low serial correlation. For example, in the $\rho = 0.2$ case, tests are only slightly over-sized when $N = 200$, with the empirical rejection probabilities falling between 0.056 and 0.082. When $N = 600$, however, the rejection probabilities are roughly twice as large, varying between 0.106 and 0.1375. With large N and ρ, the rejection probabilities approach unity. We do not observe a strong or monotone relationship between T and the rejection frequencies.

In sum, this exercise suggests the conditional logit estimator is not robust to this particular violation of CI. Neglected serial correlation in the latent model errors causes substantial bias and, accordingly, over-sized tests.

4.1.2 Unconditional Logit

While the previous subsection reported large biases for conditional logit, our experiments found uniformly higher bias in unconditional logit. Figure 1 (b) shows bias for the $N = 200$ case for different values of ρ, while Figure 2 compares the bias of conditional and unconditional logit.

As one would expect given the IPP, the finite sample bias of unconditional logit for β is substantial regardless of ρ, N, or T. For instance, in the CI case when $N = 200$, $T = 2$, we estimate the bias is 109% which is close to the theoretical finding of Abrevaya (1997).[9] As with conditional logit, bias is greater when ρ is larger or T is smaller. For example, in the experiment with $N = 200$ and $T = 5$, bias is 40% and 81% when $\rho = 0$ and $\rho = 0.6$, respectively. When $T = 2$, bias is much higher at 130% when $\rho = 0.2$ and 238% when $\rho = 0.6$. This suggests that incidental parameters and serial correlation are both potentially large sources of bias in the FE logit model.

The standard deviations of the empirical distributions of the unconditional logit slope estimator are also higher than those of conditional logit. Similar to before, we also find that increased serial correlation increases the standard deviations disproportionately when T is small. In the $N = 200$, $T = 2$ case, for example, the standard deviation is 0.58 when $\rho = 0$, but a very large 1.2 when $\rho = 0.6$ (the variances at high values of ρ fall substantially as N increases). As with conditional logit, the clustered standard errors perform well, tending to fall within 5% of the empirical standard deviations. Due to the larger biases, however, we find that over-rejection is a more severe problem with unconditional logit, regardless of ρ, N, and T. Size distortion increases with both ρ and N, and as before the relationship between rejection probability and T is non-monotone. Rejection probabilities quickly approach unity, particularly as standard errors fall with larger N. For instance, when $N = 200$ and $\rho = 0.4$, the rejection probabilities range from 0.78 to 0.91.

[Insert Figures 1 and 2 here]

[9] Our finding is qualitatively similar to Greene (2004), who with a similar design, finds bias of 102% with $N = 1,000$ over 200 replications.

4.2 Comparison of APE Estimators

4.2.1 Hybrid Conditional Logit

Fixed-T consistency of conditional logit for β under CI does not carry over to the hybrid APE, as the latter depends on estimated incidental parameters (Fernandez-Val, 2009). Therefore, Figures 3 and 4 in this section, as well as Tables 7, 13, 19, and 25 compare the uncorrected APE to one corrected using Fernandez-Val's (FV) large-T method. For all levels of ρ and N, bias in the uncorrected conditional logit APE estimates is substantial, though it declines with T and tends to be less than the bias in the conditional logit slope estimates. For instance, when $N = 200$, $\rho = 0$, the bias is -47% when $T = 2$ and -11% when $T = 8$. The bias is slightly greater for higher ρ, lower T combinations, but qualitatively similar to the $\rho = 0$ case. As expected, using the FV adjustment reduces the magnitude of the bias in all cases. For example, under the ideal condition of CI and $T \geq 6$, the bias in the corrected APE is less than 2%, while the uncorrected APE have a bias that is 10-13% percentage points higher.

The FV and related corrections reduce the order of the bias from finite-T using asymptotic expansions where T grows. Not surprisingly, then, bias in the corrected APE is larger when T is small. For instance, with $N = 200$ and $\rho = 0$, the bias is -25% when $T = 2$, but less than 2% in magnitude when $T \geq 5$. Holding T fixed, the magnitude of bias increases with ρ when T is small. For instance, in the $T = 3$ case, it grows from -9.5% when $\rho = 0$ to -13.2% when $\rho = 0.6$ and -23.7% when $\rho = 0.8$. However, changes in the bias when ρ increases are only slight when T is moderately large. When $T = 6$ and $N = 200$, for example, bias only increases from -1.8% to -2.7% when ρ increases from 0 to 0.6. When $T \geq 7$, bias is less than 2% in magnitude for all levels of ρ.

The FV correction does not affect the asymptotic variance of the estimator (Fernandez-Val, 2009). In our simulations, the empirical standard deviations of the corrected APE are slightly higher than those of the uncorrected APE. However, the differences diminish with larger T and we suspect $T = 8$ may not be large enough for this asymptotic result to hold for the variance. As a consequence, the Huber-White clustered standard errors tend to underestimate the sampling variation of the corrected APE, though the degree of underestimation diminishes as T increases.[10]

[10] In principle, one could use expansions similar to those in FV and Hahn and Newey (2004) to derive corrections for the standard errors for incidental parameters, but to our knowledge, this has not been done. The panel

For example, even with $T = 8$, the SE is too small by about 8% in the $\rho = 0$, $N = 200$ case. Moreover, neglected serial correlation exacerbates this problem, as the downward bias in the SE is about 19% when $T = 8$ and $\rho = 0.6$. As a consequence, the corrected hybrid APE produces over-sized tests. The degree of over-rejection increases with ρ, though it decreases with T as standard errors improve. For instance, despite relatively low bias in the $N = 200$, $T = 8$ case (less than 2% in magnitude), the tests of the true null hypothesis reject with probability ranging from 0.077 to 0.188. As before, larger N increases the size distortions as standard errors fall overall, with rejection probabilities ranging from 0.0830 to 0.222 in the $N = 600$, $T = 8$ case. The increase in over-rejection is even larger for smaller T.

The bottom line is, bias correction is necessary to estimate APE using slopes estimated from conditional logit, and even then, moderate to large T is required for bias to be small. While the hybrid point estimator appears relatively robust to serial correlation when T is larger, analytic inference may be misleading unless T is large. Moreover, as the next two subsections describe, lower bias and correctly-sized tests can be found with the competing estimators, at least for the DGP we consider.

4.2.2 Unconditional Logit

Some previous studies of panel binary response models have found that despite large biases in slope estimates, bias in APE estimates tends to be much smaller.[11] We find a similar pattern. For instance, in the $N = 200$ case for the longer panels ($T \geq 6$), bias is less than 2% for all levels of ρ. With zero to moderate serial correlation ($\rho \leq 0.4$), the bias is less than 2% for all $T \geq 4$. As with conditional logit, bias is greater for the shorter panels; for instance, it is -18% when $T = 2$, $N = 200$, and $\rho = 0$, and increases in magnitude to -34% when $\rho = 0.6$. The effect of serial correlation is much smaller for the longer panels, however, as even with $\rho = 0.8$, bias is less than 2% in magnitude for $T \geq 6$. As Figures 3 and 4 show, the unconditional logit APE is has uniformly lower bias than the corrected hybrid APE, in contrast with the slope estimators.[12]

Serial correlation appears to have relatively small impact on the empirical standard devia-

bootstrap is another option.

[11] See, for example, Greene (2004) or Fernandez-Val (2009).

[12] Note this "small bias" property of APE is specific to static binary response estimators and should not be expected to hold more generally.

tions, at least for larger T. However, as with the hybrid conditional logit APE, the performance of the standard errors deteriorates with higher levels of serial correlation, despite our use of clustered standard errors. For instance, even with $N = 600$, the SE's underestimate the empirical standard deviations by about 7-18% when $\rho = 0$ and by about 17-28% when $\rho = 0.6$. As a result, even with a relatively large cross section of 600, distortions in inference increase with ρ even as bias in the point estimator remains relatively constant. For instance, with $T = 8$, bias for the APE increases by only about 0.3 percentage points when we increase ρ from 0 to 0.6, but the rejection probability increases from 0.066 to 0.104. Size distortions are larger as T is smaller, as the rejection probability is estimated as 0.193 when $T = 4$, $\rho = 0.6$, driven both by a slightly larger bias and greater degree of underestimation by the SE.

4.2.3 CRE and FE OLS

As previously discussed, both the CRE logit model and the LPM are incorrect for our DGP. Nevertheless, we find they perform well for estimating the APE, and as predicted by theory, are not impacted by neglected serial correlation. In fact, these intentionally misspecified models tend to outperform both conditional and unconditional logit, especially when T is small. Since our DGP includes a single Gaussian regressor, Stoker (1986) suggests we might expect FE OLS to consistently estimate the APE. FE OLS is a good option if one is concerned with APE of continuous variables and multivariate normality of the regressors is reasonable. If binary regressors are of interest, however, CRE logit may be more attractive.

For this particular DGP, pooled logit estimation of the CRE model and FE OLS of the linear model perform very similarly quantitatively, and so we focus on the the numerical results for CRE. First of all, finite sample performance is not very sensitive to N, which we would expect if N is large enough for estimated quantities to be close to their probability limits. Moreover, even with a misspecified model of the response probability, the biases in the APE estimates are small—around -4% when $T = 2$, -2.5% when $T = 3$, and -1$ to -2% for $T = 4$ or $T = 5$. The bias is less than 1.1% in magnitude when $T \geq 6$. Most relevant relevant for this paper, the bias estimates (to two decimal places) do not change as ρ increases, even when T is small. In particular, the CRE logit APE are significantly less biased than conditional and unconditional logit when T is small and ρ is high, as shown in Figure 3 (d), and Figure 4. In contrast to conditional and unconditional logit, CRE and FE OLS do not estimate the partial effects of non-movers (units with constant outcomes over time) to be identically zero, which avoids the severe attenuation we observe in the short panels with high serial correlation.

While the biases of CRE, FE OLS, the corrected conditional logit, and unconditional logit are quantitatively similar at larger values of T, CRE and FE OLS provide more appropriately-sized tests. The clustered standard errors for CRE and FE OLS appear to better approximate the empirical standard deviations, with the difference between them less than 3% in most cases. As a result, the rejection probabilities are much closer to their nominal values, for instance ranging from 0.0510 to 0.073 when $N = 200$ and $\rho = 0.6$ (once again, we do not observe a monotone relationship between T and the rejection probability). Finally, we note that while the good performance of CRE may be linked to our underlying DGP (our equation for c_i contains a partial time-average of x_{it}, similar to the CRE model assumption), we are encouraged by the robustness of the method to small T and high ρ, in contrast to conditional and unconditional logit. Furthermore, CRE can always be made more flexible with, for example, time averages of polynomials in x_{it}, or higher functions of $\bar{x}_i$.[13]

[Insert Figures 3 and 4 here]

5 Application: Women's Welfare Participation

In this section, we illustrate the application of conditional logit and competing methods to data on the welfare participation of women. The dataset comes by way of Chay and Hyslop (2014) from the 1990 panel of the Survey of Income and Program Participation (SIPP). It contains 1,934 women over eight periods of four months in length. The sample contains women who were 18-65 years old and either received AFDC payments during or before the sample period or whose average total family income during the sample period was below the family-specific average poverty level.[14] Welfare participation status shows clear serial correlation over time. For instance, out of 1,934 women, 1,076 never participated in welfare program and 364 participated

[13]See Wooldridge and Zhu (2017) for high dimensional estimation of the CRE probit model.

[14]In summer 1993, the Nation had 36 million mothers 15 to 44 years old; 3.8 million of them (10 percent) received AFDC (Aid to Families with Dependent Children) payments supporting a total of 9.7 million children. An additional 0.5 million women over 45 years old and 0.3 million fathers living with their dependent children also received AFDC.

in all sample periods. Table 3 provides descriptive statistics for the variables used in estimation.

Table 3: Descriptive statistics

	Mean	S.D	Min	Max
AFDC partic.	.302	.459	0	1
Married	.291	.454	0	1
Kids	1.849	1.479	0	10
Black	.310	.463	0	1
Educ	11.260	2.700	0	18
Poverty level	1026.426	352.849	519.755	2377.113
Age	34.05	11.49	16	63

15,472 observations (8×1,934)

We specify the following static model for welfare participation. For a random draw i and $t = 1, 2, \ldots, T$,

$$participation_{it} = 1[\mathbf{z}_{it}\beta + \delta_t + c_i + u_{it} \geq 0], \tag{6}$$

where δ_t is a period fixed effect and $\mathbf{z}_{it}$ includes a dummy for race, years of education, family poverty level status, $age/10$, $age^2/100$, marital status and the number of children age less than 18. Assuming u_{it} is distributed logistic, we have:

$$P(participation_{it} = 1|\boldsymbol{z}_{it}, c_i, \delta_t) = \Lambda(\mathbf{z}_{it}\beta + \delta_t + c_i). \tag{7}$$

The covariates of interest in our application are marital status (*married*) and the number of children (*kids*). Our specification differs from Chay and Hyslop, who also include lagged participation to model what they refer to as "structural state dependence," in addition to time-constant individual heterogeneity. We abstract from this issue, as conditional logit requires strict exogeneity, but researchers may also be interested in the average marginal effect of marriage or children on the probability of participation without holding constant past participation. Of course, if past values of participation belongs in the logit function in Eq. (7), then our model for $P(participation_{it} = 1|\boldsymbol{z}_{it}, c_i, \delta_t)$ is incorrect.

Table 4: Estimation results: women's welfare participation

	(1)		(2)		(3)		(4)	(5)		(6)
Estimator	Cond. Logit*		Uncond. Logit		CRE Logit		FE OLS	CRE Logit**		FE OLS**
Parameter	Slope	APE	Slope	APE	Slope	APE	APE	Slope	APE	APE
Married	-2.802	-0.113	-3.432	-0.114	-1.801	-0.262	-0.271	-2.508	-0.112	-0.126
	(0.303)	(0.008)	(0.410)	(0.007)	(0.196)	(0.023)	(0.028)	(0.268)	(0.007)	(0.011)
Kids	0.761	0.036	0.895	0.037	0.286	0.046	0.052	0.614	0.035	0.033
	(0.156)	(0.007)	(0.185)	(0.007)	(0.057)	(0.009)	(0.010)	(0.131)	(0.007)	(0.007)
Marriedbar					-0.560			2.070		
					(0.235)			(0.294)		
Kidsbar					0.302			-0.288		
					(0.087)			(0.143)		
Ratio	-3.684		-3.835		-6.300		-5.171	-4.082		-3.841
	(0.856)		(0.907)		(1.461)		(1.145)	(0.981)		(0.913)
LL	-1412.6215		-1940.9004		-7551.0719			-2505.0202		
Sample Size	494		494		1934		1934	494		494

*APE estimated using hybrid approach with FV correction as described in Section 2.1

**Estimated using only observations that change participation in the sample period. To mimic conditional and unconditional logit, non-changing observations assigned a partial effect of zero.

Table 4 gives slope and APE estimates across different model specifications and estimators. Columns 1, 2, 3, and 4 correspond to methods tested in our Monte Carlo experiment. The conditional logit slopes are smaller in magnitude than the unconditional logit slopes, which is to be expected given the latter suffers from the IPP. The simulation results suggest that these estimates are likely to be significantly biased if there is neglected serial correlation in the participation variable, as seems likely. Slope estimates between CRE logit and the other two logit models are not comparable because of scaling.

Our simulation results also suggest that with a longer panel (i.e. $T = 8$ as in this case) the hybrid conditional logit approach, unconditional logit, and CRE logit will yield similar results for the APE, across different degrees of serial correlation. This does not appear to be the case with this data set. For example, the CRE logit estimates imply that holding kids and individual characteristics constant, marriage is associated with a 26.2 percentage point decline in the probability of receiving AFDC benefits, while the hybrid conditional logit estimate is for only a 11.3 percentage point decline (11.2 for unconditional logit). The APE of kids is also smaller in magnitude, though not so drastically. One key difference between this dataset and our simulated panels, however, is only 25% of the observations in the AFDC dataset change participation status during the sample period. This proportion, as well as the wedge between the full-sample estimates (CRE and FE OLS) and the reduced sample estimates (conditional and unconditional logit) is more closely matched in the simulations with the short panels with high serial correlation. The low proportion of "movers" in the AFDC data could reflect additional persistence in participation stemming from serial correlation in marriage status and number of kids, as well as possibly larger individual heterogeneity.

In our simulations, increased persistence caused decreases in the proportion of the sample that had changing outcomes over time (see Table 2). This was associated with larger bias in conditional and unconditional logit APE. One reason for this bias is that these estimators implicitly treat the never-movers as having a participation probability of zero. This attenuates the APE estimates, as we would not expect their chance of participation to be identically zero.[15] To see the extent of this attenuation bias, we repeat the CRE Logit and FE OLS estimation in

[15] Recall that the MLE of the fixed effects are unbounded when the response variable for an individual is constant over time, and so the participation probability, which is the scale factor in the APE formula, is zero in the limit.

columns 5 and 6, but only on the subsample of women with changing participation. To mimic the unconditional and conditional logit, we calculate the APE treating the never movers as having a partial effect that is identically zero. The resulting APE are attenuated to nearly the same level as the unconditional and hybrid conditional logit estimates, suggesting this source of bias is quite important.[16] These results, as well as our simulation findings, therefore, suggest that the CRE or LPM approach may be preferred to either conditional or unconditional logit in this case.

6 Conclusion

This paper examines the robustness of the conditional logit estimator for the binary response panel data model with unobserved effects. Our primary interest is in how the parameter estimates behave under violations of the conditional independence assumption. We do so by generating serially correlated latent errors that nevertheless have marginal logistic distributions. Simulation results show that conditional logit method is not robust to violation of CI assumption, even for rather simple deviations from CI. As expected, the magnitude of bias in the coefficient is grows as the time dimension shrinks or as the serial correlation gets higher. We also examine the bias of the unconditional logit estimator and find that, in addition to the well-known bias cased by the incidental parameters problem, serial correlation is also a substantial source of bias. These findings add to our understanding of the tradeoffs among different estimation approaches.

We also studied the properties of three estimators of average partial effects. When T is larger, a hybrid conditional logit estimator of the APE has small bias, but standard errors tend to be too small. Both CRE logit and FE on the linear model appear robust to high serial correlation in the errors, though the results may be somewhat dependent on the specific DGP we chose.

The findings we report in this paper imply that the conditional logit approach cannot be uniformly preferred to strategies such as correlated random effects, as the two approaches produce consistent parameters (or scaled parameters) under nonoverlapping sets of assumptions. Given that conditional independence is a strong assumption, we may not always feel comfortable with conditional logit, even when we are satisfied with parameter estimators. When we want to

[16]Dropping the never movers from the average entirely is another option, though this results in the APE for the subpopulation of movers, which is higher than the unconditional APE in this example.

estimate APEs, the CL approach poses additional challenges. The hybrid approach we study here can work reasonably well, but not across all scenarios with small T and/or serial correlation.

References

Abrevaya, J. (1997). “The equivalence of two estimators of the fixed-effects logit model”. In: *Economics Letters* 55, pp. 41–43.

Alexander, B. and R. Breunig (2014). “A Monte Carlo study of bias corrections for panel probit models”. In: *Journal of Statistical Computation and Simulation* 86.1, pp. 74–90.

Andersen, E. B. (1970). “Asymptotic Properties of Conditional Maximum-Likelihood Estimators”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 32.2, pp. 283–301. ISSN: 00359246. DOI: 10.2307/2984535.

— (1973). *Conditional Inference and Models for Measuring.* Copenhague: Mentalhygienjnisk Forlag.

Arellano, M. (2003). “Discrete Choices with Panel Data”. In: *Investigaciones Economicas* 27, pp. 423–458.

Arellano, M. and J. Hahn (2007). *Understanding Bias in Nonlinear Panel Models: Some Recent Developments. In Advances in Economics and Econometrics, Blundell R, Newey W, Persson T (eds).* Cambridge: Cambridge University Press.

Chamberlain, G. (1980). “Analysis of Covariance with Qualitative Data”. In: *Review of Economic Studies* 47, pp. 225–238.

— (1984). “”Panel Data”, in eds Z. Griliches and M. Intriligator”. In: *Handbook of Econometrics* 2, pp. 1248–1313.

Chay, K. and D. Hyslop (2014). “Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence Using Alternative Approaches”. In: *Safety Nets and Benefit Dependence (Research in Labor Economics)*, pp. 1–39.

Collado, M. and M. Browning (2007). “Habits and heterogeneity in demands: a panel data analysis”. In: *Journal of Applied Econometrics* 22, pp. 625–640.

Devroye, L. (1986). *Non-Uniform Random Variate Generation.* New York: Springer-Verlag.

Fernandez-Val, I. (2009). "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models". In: *Journal of Econometrics* 150, pp. 71–85.

Greene, WH. (2004). "The Behavior of the Fixed Effects Estimator in Nonlinear Models". In: *The Econometrics Journal* 7, pp. 98–119.

— (2012). *Econometric analysis*. Prentice Hall.

Hahn, J. and W. Newey (2004). "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models". In: *Econometrica* 72, pp. 1295–1319.

Hajivassiliou, V. and D. McFadden (1998). "The Method of Simulated Scores with Application to Models of External Debt Crises". In: *Econometrica* 66, pp. 863–896.

Heckman, J. (1981). *The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Descrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence*. Cambridge, Massachusetts: MIT Press.

Johnson, M. (1987). *Multivariate Statistical Simulation*. New York: John Wiley and Sons.

Lancaster, T. (2000). "The Incidental Parameters Problem since 1948". In: *Journal of Econometrics* 95, pp. 391–413.

Neyman, J. and E. Scott (1948). "Consistent Estimates Based on Partially Consistent Observations". In: *Econometrica* 16, pp. 1–32.

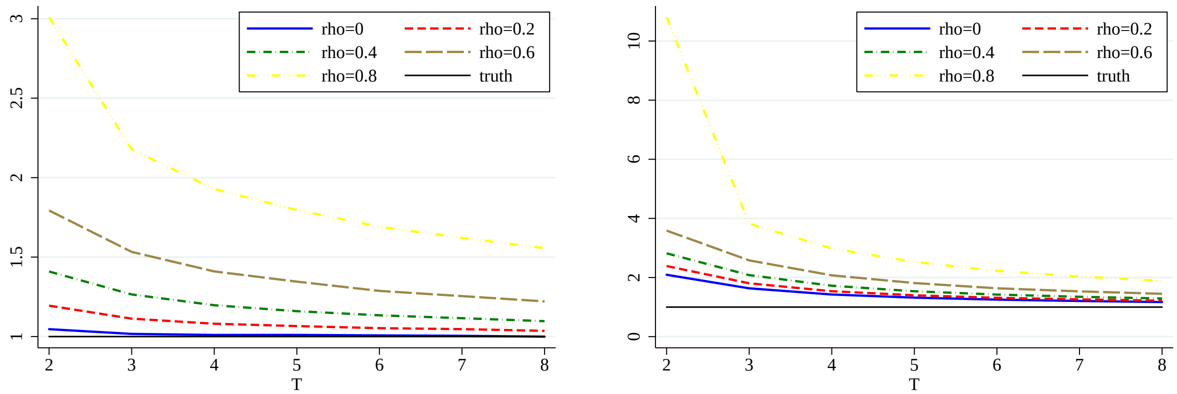
Stoker, TM. (1986). "Consistent Estimation of Scaled Coefficients". In: *Econometrica* 54.6, p. 1461.

Wooldridge, J.M. (1999). "Distribution-free estimation of some nonlinear panel data models". In: *Journal of Econometrics* 90.1, pp. 77–97. DOI: 10.1016/s0304-4076(98)00033-5.

— (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Wooldridge, J.M. and Y. Zhu (2017). "L1-Regularized Quasi-Maximum Likelihood Estimation and Inference in High-Dimensional Correlated Random Effects Probit". In: *Journal of Business & Econ. Statistics*. Forthcoming.

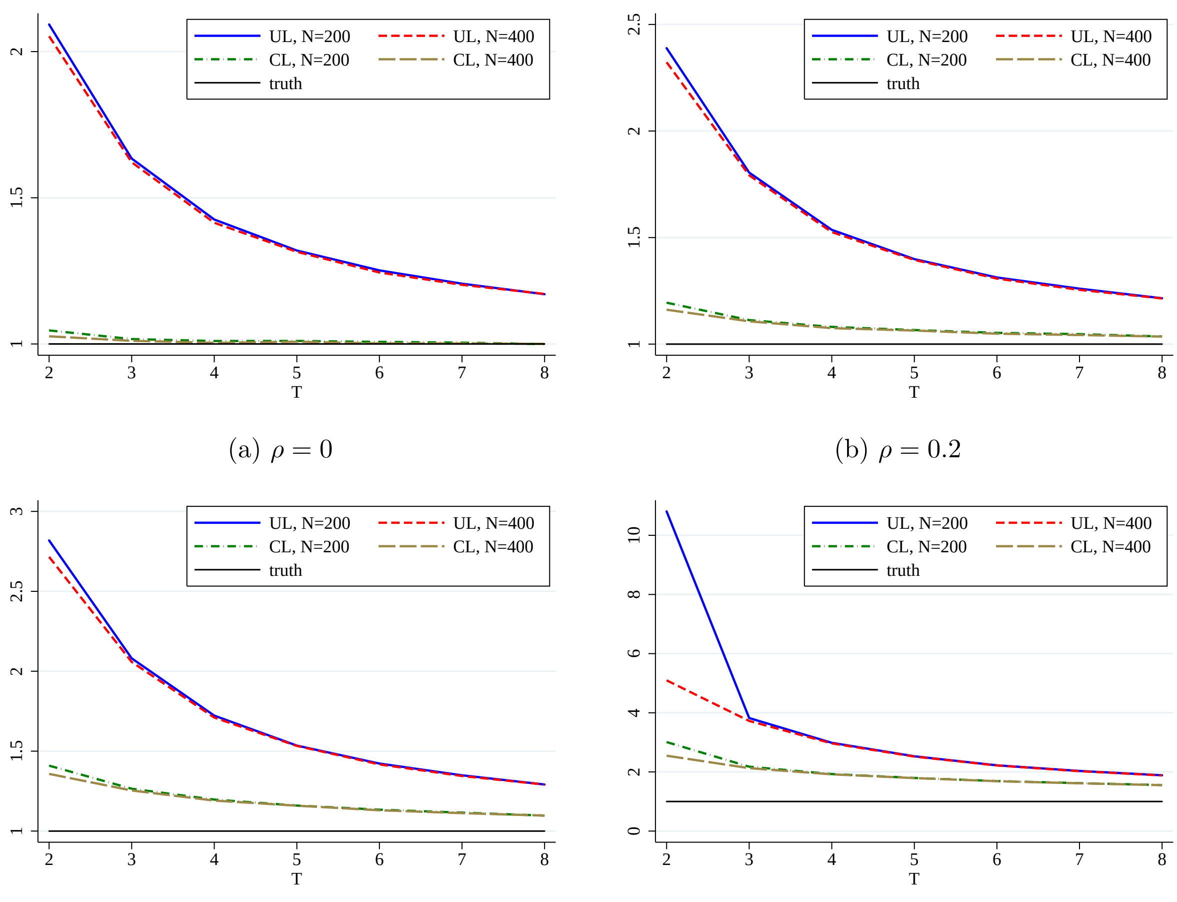
Figure 1: Empirical means for estimators of β, $N = 200$



(a) Conditional Logit

(b) Unconditional Logit

Figure 2: Comparing uncond. and cond. logit estimators of β



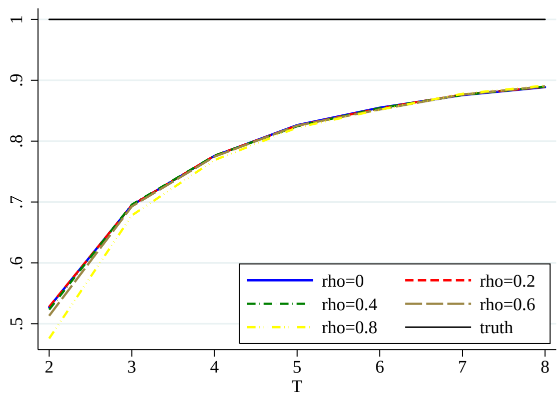
(a) $\rho = 0$

(b) $\rho = 0.2$

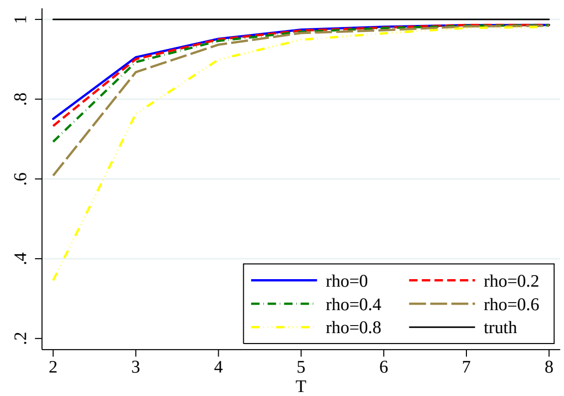
(c) $\rho = 0.6$

(d) $\rho = 0.8$

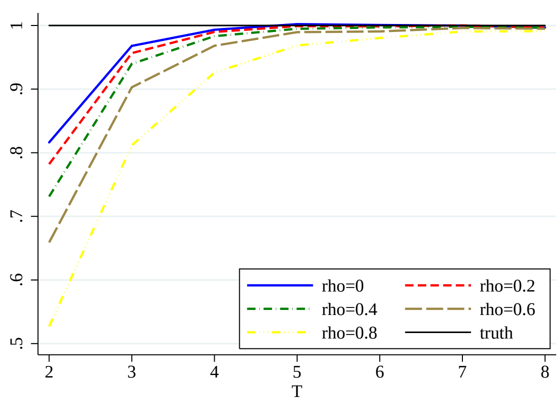
Figure 3: Empirical means for estimators of APE_x, $N = 200$



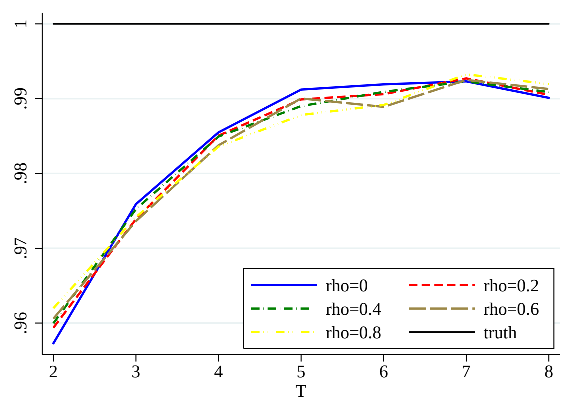
(a) Hybrid Conditional Logit - Uncorrected



(b) Hybrid Conditional Logit - Corrected

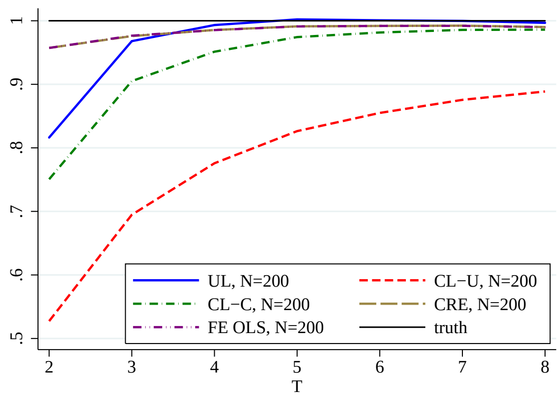


(c) Unconditional Logit

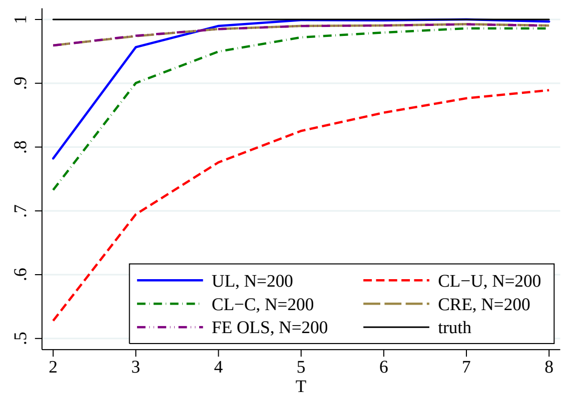


(d) Correlated Random Effects

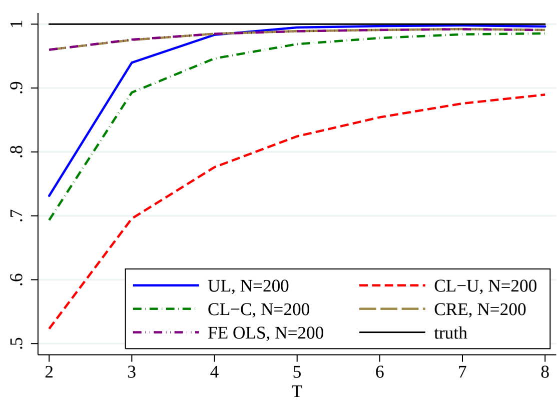
Figure 4: Comparing estimators of APE_x



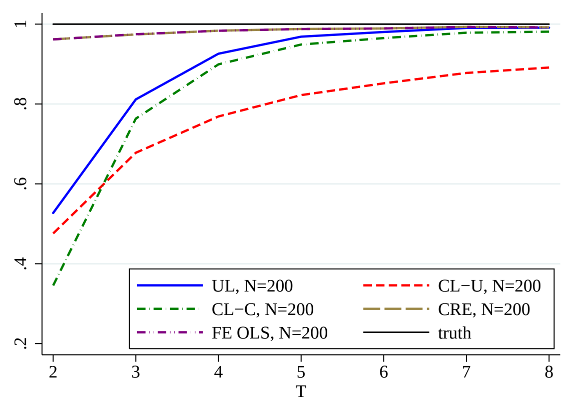
(a) $\rho = 0$



(b) $\rho = 0.2$



(c) $\rho = 0.4$



(d) $\rho = 0.8$

Appendix: Monte Carlo Simulation Results

Table 5: Conditional Logit Estimates of β (true value=1), $N = 100$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	1.1162	0.4831	0.8100	0.0560
T=3	1.0187	0.2475	0.9586	0.0540
T=4	1.0270	0.1876	1.0006	0.0470
T=5	1.0105	0.1621	0.9660	0.0540
T=6	1.0078	0.1409	0.9788	0.0540
T=7	1.0078	0.1266	0.9866	0.0525
T=8	1.0059	0.1178	0.9735	0.0590
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	1.2933	0.6519	0.6997	0.0555
T=3	1.1179	0.2705	0.9565	0.0520
T=4	1.0967	0.2060	0.9667	0.0600
T=5	1.0680	0.1678	0.9801	0.0600
T=6	1.0554	0.1429	1.0037	0.0540
T=7	1.0479	0.1322	0.9766	0.0630
T=8	1.0413	0.1210	0.9750	0.0605
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	1.5409	0.7765	0.7113	0.0705
T=3	1.2773	0.3115	0.9489	0.0885
T=4	1.2085	0.2272	0.9630	0.1055
T=5	1.1674	0.1842	0.9672	0.1290
T=6	1.1342	0.1566	0.9807	0.1175
T=7	1.1200	0.1422	0.9611	0.1275
T=8	1.1037	0.1277	0.9730	0.1180
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	2.3937	10.0210	0.0789	0.1428
T=3	1.5725	0.4233	0.8738	0.2450
T=4	1.4352	0.2744	0.9521	0.3270
T=5	1.3463	0.2155	0.9509	0.3490
T=6	1.2873	0.1738	1.0012	0.3485
T=7	1.2586	0.1573	0.9757	0.3575
T=8	1.2304	0.1408	0.9812	0.3580
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	5.0688	22.0286	0.0727	0.2446
T=3	2.3103	0.8878	0.6728	0.6035
T=4	1.9883	0.4257	0.8994	0.8150
T=5	1.8158	0.2960	0.9613	0.8885
T=6	1.7014	0.2449	0.9563	0.8970
T=7	1.6315	0.2047	0.9820	0.9315
T=8	1.5639	0.1773	1.0017	0.9240

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 6: Unconditional Logit Estimates of β (true value=1), $N = 100$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	2.2324	0.9662	0.8100	0.2575
T=3	1.6403	0.4336	0.9543	0.2630
T=4	1.4529	0.2898	0.9950	0.2865
T=5	1.3215	0.2271	0.9648	0.2760
T=6	1.2532	0.1855	0.9774	0.2480
T=7	1.2107	0.1594	0.9866	0.2405
T=8	1.1782	0.1436	0.9738	0.2225
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	2.5865	1.3039	0.6997	0.3110
T=3	1.8173	0.4811	0.9523	0.3850
T=4	1.5632	0.3229	0.9616	0.3940
T=5	1.4033	0.2377	0.9782	0.3755
T=6	1.3170	0.1899	1.0018	0.3475
T=7	1.2622	0.1676	0.9756	0.3320
T=8	1.2222	0.1483	0.9752	0.3155
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	3.2870	9.3014	0.1187	0.3955
T=3	2.1069	0.5667	0.9432	0.5335
T=4	1.7432	0.3654	0.9548	0.5670
T=5	1.5480	0.2665	0.9641	0.5590
T=6	1.4242	0.2110	0.9792	0.5175
T=7	1.3556	0.1826	0.9595	0.5155
T=8	1.3001	0.1579	0.9736	0.4905
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	21.9164	281.1146	0.0055	0.5095
T=3	2.6585	0.7961	0.8655	0.7520
T=4	2.1199	0.4586	0.9461	0.8175
T=5	1.8161	0.3234	0.9456	0.8160
T=6	1.6366	0.2402	1.0014	0.7950
T=7	1.5380	0.2065	0.9742	0.8035
T=8	1.4605	0.1776	0.9799	0.7790
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	103.1982	968.1338	0.0027	0.6428
T=3	4.0864	1.7472	0.6619	0.9130
T=4	3.0974	0.7651	0.8940	0.9880
T=5	2.5635	0.4794	0.9605	0.9910
T=6	2.2417	0.3646	0.9541	0.9935
T=7	2.0505	0.2854	0.9803	0.9930
T=8	1.8968	0.2349	1.0002	0.9930

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 7: Hybrid* Conditional Logit Estimates of APE_x (true value = 1), $N = 100$

	Uncorrected				With FV Correction			
$\rho = 0.0$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5410	0.1747	0.9633	0.8124	0.7496	0.2113	0.7961	0.3632
T=3	0.6908	0.1420	1.0360	0.5710	0.8985	0.1771	0.8308	0.1510
T=4	0.7831	0.1232	1.0240	0.4065	0.9596	0.1476	0.8544	0.1160
T=5	0.8246	0.1147	0.9695	0.3440	0.9721	0.1321	0.8418	0.1150
T=6	0.8544	0.1031	0.9738	0.3135	0.9806	0.1160	0.8653	0.0915
T=7	0.8764	0.0955	0.9680	0.2720	0.9864	0.1053	0.8779	0.0890
T=8	0.8923	0.0895	0.9642	0.2425	0.9899	0.0975	0.8857	0.0870
$\rho = 0.2$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5387	0.1675	0.8709	0.8701	0.7187	0.2142	0.6810	0.5065
T=3	0.6925	0.1394	0.9750	0.6335	0.8956	0.1713	0.7931	0.1890
T=4	0.7817	0.1237	0.9694	0.4590	0.9560	0.1474	0.8134	0.1305
T=5	0.8256	0.1111	0.9646	0.3855	0.9719	0.1278	0.8384	0.1090
T=6	0.8545	0.1008	0.9677	0.3345	0.9799	0.1132	0.8612	0.0955
T=7	0.8758	0.0956	0.9433	0.2970	0.9851	0.1053	0.8564	0.0910
T=8	0.8926	0.0888	0.9529	0.2525	0.9896	0.0967	0.8757	0.0910
$\rho = 0.4$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5331	0.1578	0.7670	0.9221	0.6677	0.2028	0.5969	0.7035
T=3	0.6949	0.1369	0.8796	0.7140	0.8885	0.1635	0.7364	0.2505
T=4	0.7786	0.1215	0.9136	0.5405	0.9488	0.1438	0.7716	0.1620
T=5	0.8269	0.1121	0.8991	0.4245	0.9713	0.1287	0.7830	0.1380
T=6	0.8532	0.1020	0.9145	0.3655	0.9768	0.1148	0.8131	0.1180
T=7	0.8766	0.0963	0.8996	0.3205	0.9846	0.1060	0.8178	0.1130
T=8	0.8937	0.0887	0.9259	0.2780	0.9897	0.0965	0.8511	0.0905
$\rho = 0.6$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5251	0.1466	0.6722	0.9396	0.5425	0.3057	0.3225	0.8874
T=3	0.6969	0.1339	0.7433	0.7945	0.8639	0.1513	0.6578	0.3895
T=4	0.7799	0.1217	0.7936	0.6260	0.9415	0.1418	0.6814	0.2355
T=5	0.8249	0.1130	0.8044	0.5065	0.9646	0.1293	0.7033	0.1890
T=6	0.8509	0.1012	0.8516	0.4340	0.9716	0.1138	0.7571	0.1580
T=7	0.8763	0.0967	0.8410	0.3610	0.9821	0.1065	0.7634	0.1500
T=8	0.8951	0.0886	0.8790	0.3035	0.9893	0.0963	0.8082	0.1270
$\rho = 0.8$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.4831	0.1311	0.6581	0.9620	0.2545	0.4730	0.1824	0.9881
T=3	0.6843	0.1285	0.5973	0.8749	0.7387	0.1762	0.4354	0.7543
T=4	0.7755	0.1239	0.6099	0.7402	0.8996	0.1383	0.5468	0.4209
T=5	0.8238	0.1123	0.6576	0.6260	0.9486	0.1266	0.5833	0.3015
T=6	0.8524	0.1059	0.6847	0.5380	0.9655	0.1189	0.6102	0.2625
T=7	0.8776	0.0977	0.7193	0.4650	0.9782	0.1073	0.6549	0.2070
T=8	0.8957	0.0905	0.7642	0.3810	0.9859	0.0985	0.7021	0.1810

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors. Heterogeneity is estimated by unconditional MLE with slopes restricted to their conditional logit estimates.

Table 8: Unconditional Logit Estimates of APE_x (true value = 1), $N = 100$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.8204	0.2062	0.8101	0.2740
T=3	0.9599	0.1782	0.8775	0.0960
T=4	1.0005	0.1491	0.8995	0.0850
T=5	0.9996	0.1332	0.8719	0.0935
T=6	0.9998	0.1164	0.8903	0.0785
T=7	1.0004	0.1054	0.8979	0.0810
T=8	1.0003	0.0975	0.9027	0.0785
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.7845	0.1926	0.7906	0.3975
T=3	0.9509	0.1722	0.8486	0.1075
T=4	0.9950	0.1485	0.8675	0.0935
T=5	0.9988	0.1286	0.8768	0.0795
T=6	0.9988	0.1136	0.8903	0.0770
T=7	0.9990	0.1055	0.8784	0.0790
T=8	1.0002	0.0967	0.8946	0.0815
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.7339	0.1831	0.7561	0.5390
T=3	0.9353	0.1650	0.8063	0.1530
T=4	0.9849	0.1448	0.8373	0.1105
T=5	0.9970	0.1293	0.8302	0.1050
T=6	0.9955	0.1150	0.8489	0.0960
T=7	0.9986	0.1061	0.8448	0.1005
T=8	1.0005	0.0965	0.8741	0.0780
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.6574	0.1960	0.7075	0.6740
T=3	0.9030	0.1567	0.7532	0.2570
T=4	0.9725	0.1434	0.7669	0.1495
T=5	0.9883	0.1299	0.7664	0.1385
T=6	0.9892	0.1139	0.8058	0.1195
T=7	0.9957	0.1064	0.8013	0.1215
T=8	1.0001	0.0963	0.8390	0.1045
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.5105	0.2805	0.5835	0.7924
T=3	0.8139	0.1552	0.7406	0.4635
T=4	0.9306	0.1463	0.6735	0.2685
T=5	0.9695	0.1288	0.6884	0.1965
T=6	0.9806	0.1192	0.6923	0.1880
T=7	0.9902	0.1073	0.7196	0.1580
T=8	0.9955	0.0982	0.7546	0.1470

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 9: CRE Logit Estimates of APE_x (true value = 1), $N = 100$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9605	0.2585	0.9631	0.0685
T=3	0.9640	0.1767	0.9867	0.0680
T=4	0.9908	0.1405	1.0185	0.0495
T=5	0.9886	0.1269	0.9770	0.0590
T=6	0.9903	0.1094	1.0162	0.0490
T=7	0.9922	0.1026	0.9930	0.0545
T=8	0.9937	0.0950	0.9980	0.0580
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9612	0.2483	0.9652	0.0700
T=3	0.9648	0.1727	0.9853	0.0665
T=4	0.9883	0.1387	1.0150	0.0490
T=5	0.9895	0.1241	0.9882	0.0535
T=6	0.9906	0.1069	1.0330	0.0430
T=7	0.9916	0.1028	0.9856	0.0535
T=8	0.9938	0.0942	1.0012	0.0550
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9619	0.2389	0.9623	0.0700
T=3	0.9677	0.1687	0.9790	0.0665
T=4	0.9842	0.1348	1.0253	0.0465
T=5	0.9907	0.1234	0.9810	0.0570
T=6	0.9879	0.1080	1.0140	0.0495
T=7	0.9924	0.1041	0.9661	0.0570
T=8	0.9951	0.0942	0.9974	0.0530
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9641	0.2275	0.9650	0.0740
T=3	0.9717	0.1638	0.9765	0.0690
T=4	0.9862	0.1366	0.9896	0.0575
T=5	0.9889	0.1224	0.9755	0.0625
T=6	0.9851	0.1070	1.0163	0.0550
T=7	0.9920	0.1036	0.9720	0.0560
T=8	0.9967	0.0943	1.0000	0.0475
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9622	0.2169	0.9703	0.0730
T=3	0.9718	0.1578	0.9854	0.0700
T=4	0.9853	0.1341	0.9941	0.0545
T=5	0.9892	0.1190	0.9992	0.0570
T=6	0.9871	0.1101	0.9940	0.0605
T=7	0.9931	0.1039	0.9817	0.0520
T=8	0.9959	0.0961	1.0030	0.0520

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 10: FE OLS (LPM) Estimates of APE_x (true value = 1), $N = 100$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9613	0.2593	0.9664	0.0685
T=3	0.9644	0.1775	0.9859	0.0655
T=4	0.9914	0.1411	1.0187	0.0495
T=5	0.9887	0.1272	0.9777	0.0565
T=6	0.9903	0.1100	1.0141	0.0470
T=7	0.9925	0.1029	0.9940	0.0580
T=8	0.9936	0.0950	1.0004	0.0585
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9617	0.2491	0.9685	0.0690
T=3	0.9651	0.1731	0.9868	0.0640
T=4	0.9890	0.1394	1.0148	0.0470
T=5	0.9896	0.1243	0.9899	0.0520
T=6	0.9906	0.1076	1.0296	0.0425
T=7	0.9919	0.1031	0.9861	0.0545
T=8	0.9936	0.0943	1.0035	0.0545
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9624	0.2395	0.9658	0.0695
T=3	0.9679	0.1690	0.9810	0.0645
T=4	0.9847	0.1355	1.0242	0.0490
T=5	0.9908	0.1237	0.9819	0.0575
T=6	0.9879	0.1084	1.0132	0.0495
T=7	0.9927	0.1044	0.9665	0.0565
T=8	0.9950	0.0944	0.9993	0.0535
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9646	0.2277	0.9706	0.0745
T=3	0.9720	0.1644	0.9761	0.0695
T=4	0.9868	0.1372	0.9893	0.0600
T=5	0.9889	0.1228	0.9764	0.0635
T=6	0.9850	0.1073	1.0168	0.0535
T=7	0.9922	0.1038	0.9729	0.0565
T=8	0.9965	0.0944	1.0020	0.0475
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9628	0.2172	0.9748	0.0730
T=3	0.9719	0.1588	0.9820	0.0695
T=4	0.9859	0.1347	0.9941	0.0540
T=5	0.9892	0.1196	0.9976	0.0595
T=6	0.9870	0.1102	0.9964	0.0595
T=7	0.9934	0.1043	0.9817	0.0535
T=8	0.9958	0.0963	1.0051	0.0535

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 11: Conditional Logit Estimates of β (true value=1), $N = 200$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	1.0464	0.2915	0.9046	0.0570
T=3	1.0168	0.1693	0.9847	0.0450
T=4	1.0101	0.1368	0.9530	0.0615
T=5	1.0104	0.1084	1.0188	0.0390
T=6	1.0073	0.0991	0.9849	0.0570
T=7	1.0046	0.0889	0.9898	0.0525
T=8	0.9996	0.0801	1.0096	0.0485
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	1.1943	0.3366	0.8953	0.0560
T=3	1.1127	0.1897	0.9569	0.0650
T=4	1.0807	0.1448	0.9554	0.0820
T=5	1.0660	0.1135	1.0205	0.0680
T=6	1.0528	0.1010	1.0035	0.0700
T=7	1.0468	0.0918	0.9920	0.0750
T=8	1.0359	0.0830	1.0039	0.0575
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	1.4093	0.4057	0.8919	0.1095
T=3	1.2650	0.2122	0.9716	0.1850
T=4	1.1971	0.1590	0.9585	0.2130
T=5	1.1597	0.1240	1.0101	0.2175
T=6	1.1339	0.1080	1.0038	0.2125
T=7	1.1155	0.0976	0.9882	0.2045
T=8	1.0970	0.0868	1.0109	0.1715
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	1.7933	0.5995	0.8038	0.2620
T=3	1.5329	0.2685	0.9477	0.5340
T=4	1.4099	0.1865	0.9676	0.6440
T=5	1.3457	0.1457	0.9921	0.6925
T=6	1.2875	0.1258	0.9753	0.6520
T=7	1.2546	0.1077	1.0035	0.6630
T=8	1.2214	0.0980	0.9917	0.6250
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	3.0099	4.7289	0.1887	0.5665
T=3	2.1773	0.4334	0.9130	0.9540
T=4	1.9286	0.2696	0.9689	0.9900
T=5	1.7964	0.2012	0.9911	0.9975
T=6	1.6912	0.1641	1.0019	0.9990
T=7	1.6212	0.1441	0.9831	0.9985
T=8	1.5565	0.1236	1.0114	0.9985

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 12: Unconditional Logit Estimates of β (true value=1), $N = 200$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	2.0928	0.5830	0.9046	0.5250
T=3	1.6342	0.2958	0.9819	0.5910
T=4	1.4257	0.2094	0.9530	0.5585
T=5	1.3196	0.1517	1.0173	0.5260
T=6	1.2517	0.1304	0.9835	0.4880
T=7	1.2063	0.1119	0.9896	0.4420
T=8	1.1702	0.0978	1.0079	0.3875
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	2.3887	0.6732	0.8953	0.6550
T=3	1.8047	0.3362	0.9543	0.7425
T=4	1.5363	0.2248	0.9556	0.7215
T=5	1.3989	0.1606	1.0183	0.7070
T=6	1.3127	0.1339	1.0028	0.6625
T=7	1.2602	0.1163	0.9913	0.6205
T=8	1.2153	0.1017	1.0037	0.5635
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	2.8186	0.8114	0.8919	0.8005
T=3	2.0808	0.3833	0.9702	0.9080
T=4	1.7224	0.2524	0.9589	0.8975
T=5	1.5349	0.1789	1.0070	0.8875
T=6	1.4226	0.1453	1.0023	0.8680
T=7	1.3490	0.1249	0.9877	0.8350
T=8	1.2914	0.1073	1.0109	0.7780
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	3.5867	1.1990	0.8038	0.9035
T=3	2.5796	0.4990	0.9462	0.9890
T=4	2.0738	0.3078	0.9678	0.9905
T=5	1.8127	0.2184	0.9877	0.9930
T=6	1.6354	0.1738	0.9747	0.9830
T=7	1.5316	0.1412	1.0019	0.9840
T=8	1.4486	0.1232	0.9920	0.9810
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	10.8084	75.8359	0.0232	0.9410
T=3	3.8222	0.8344	0.9140	0.9995
T=4	2.9852	0.4795	0.9662	1.0000
T=5	2.5290	0.3271	0.9838	1.0000
T=6	2.2241	0.2432	1.0013	1.0000
T=7	2.0336	0.2003	0.9808	1.0000
T=8	1.8866	0.1628	1.0136	1.0000

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 13: Hybrid* Conditional Logit Estimates of APE_x (true value = 1), $N = 200$

	Uncorrected				With FV Correction			
$\rho = 0.0$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5274	0.1216	0.9750	0.9825	0.7506	0.1517	0.7816	0.5590
T=3	0.6947	0.1007	1.0281	0.8580	0.9051	0.1260	0.8210	0.2115
T=4	0.7758	0.0909	0.9803	0.7160	0.9513	0.1089	0.8184	0.1440
T=5	0.8263	0.0768	1.0222	0.6010	0.9743	0.0885	0.8869	0.0930
T=6	0.8550	0.0738	0.9638	0.5295	0.9815	0.0829	0.8573	0.1010
T=7	0.8756	0.0684	0.9542	0.4835	0.9857	0.0755	0.8656	0.0950
T=8	0.8887	0.0612	0.9988	0.4490	0.9860	0.0665	0.9188	0.0770
$\rho = 0.2$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5278	0.1170	0.8678	0.9880	0.7328	0.1359	0.7471	0.7045
T=3	0.6946	0.0997	0.9585	0.8860	0.9004	0.1230	0.7769	0.2500
T=4	0.7761	0.0897	0.9426	0.7505	0.9498	0.1071	0.7896	0.1720
T=5	0.8254	0.0765	0.9879	0.6295	0.9720	0.0882	0.8572	0.1115
T=6	0.8540	0.0723	0.9546	0.5715	0.9796	0.0813	0.8496	0.1065
T=7	0.8765	0.0677	0.9409	0.4930	0.9861	0.0747	0.8528	0.0970
T=8	0.8892	0.0613	0.9779	0.4615	0.9860	0.0666	0.8999	0.0840
$\rho = 0.4$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5233	0.1104	0.7496	0.9935	0.6933	0.1177	0.7034	0.8813
T=3	0.6957	0.0965	0.8772	0.9270	0.8930	0.1161	0.7285	0.3235
T=4	0.7761	0.0882	0.8838	0.7900	0.9462	0.1047	0.7450	0.2110
T=5	0.8244	0.0769	0.9259	0.6680	0.9688	0.0885	0.8047	0.1460
T=6	0.8542	0.0716	0.9201	0.5980	0.9783	0.0804	0.8189	0.1160
T=7	0.8758	0.0664	0.9242	0.5275	0.9840	0.0732	0.8383	0.1075
T=8	0.8896	0.0609	0.9553	0.4780	0.9854	0.0662	0.8791	0.0875
$\rho = 0.6$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5131	0.1006	0.6327	0.9985	0.6084	0.1264	0.5035	0.9787
T=3	0.6928	0.0932	0.7472	0.9615	0.8678	0.1059	0.6571	0.5095
T=4	0.7745	0.0870	0.7823	0.8490	0.9366	0.1019	0.6683	0.2890
T=5	0.8255	0.0759	0.8430	0.7305	0.9659	0.0871	0.7351	0.1875
T=6	0.8520	0.0726	0.8357	0.6625	0.9729	0.0815	0.7443	0.1670
T=7	0.8764	0.0668	0.8597	0.5570	0.9824	0.0737	0.7797	0.1350
T=8	0.8901	0.0625	0.8796	0.5185	0.9840	0.0680	0.8089	0.1195
$\rho = 0.8$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.4758	0.0882	0.5958	0.9995	0.3455	0.2756	0.1906	1.0000
T=3	0.6779	0.0877	0.5865	0.9855	0.7634	0.1040	0.4946	0.8897
T=4	0.7688	0.0842	0.6271	0.9285	0.8993	0.0948	0.5573	0.5083
T=5	0.8223	0.0770	0.6772	0.8205	0.9489	0.0872	0.5978	0.3270
T=6	0.8517	0.0728	0.7026	0.7465	0.9651	0.0813	0.6289	0.2715
T=7	0.8778	0.0680	0.7324	0.6440	0.9785	0.0748	0.6655	0.2220
T=8	0.8913	0.0639	0.7622	0.5910	0.9812	0.0696	0.6996	0.1880

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors. Heterogeneity is estimated by unconditional MLE with slopes restricted to their conditional logit estimates.

Table 14: Unconditional Logit Estimates of APE_x (true value = 1), $N = 200$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.8163	0.1429	0.8040	0.4045
T=3	0.9679	0.1261	0.8729	0.0925
T=4	0.9933	0.1098	0.8606	0.0945
T=5	1.0021	0.0892	0.9186	0.0685
T=6	1.0008	0.0833	0.8815	0.0820
T=7	0.9998	0.0756	0.8849	0.0855
T=8	0.9967	0.0666	0.9361	0.0690
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.7821	0.1320	0.7733	0.5720
T=3	0.9566	0.1223	0.8405	0.1225
T=4	0.9899	0.1077	0.8403	0.1065
T=5	0.9991	0.0888	0.8947	0.0815
T=6	0.9986	0.0815	0.8784	0.0820
T=7	1.0001	0.0748	0.8755	0.0855
T=8	0.9968	0.0666	0.9197	0.0730
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.7313	0.1219	0.7460	0.7530
T=3	0.9397	0.1157	0.8046	0.1795
T=4	0.9832	0.1051	0.8089	0.1195
T=5	0.9948	0.0890	0.8513	0.1050
T=6	0.9970	0.0806	0.8546	0.0925
T=7	0.9980	0.0732	0.8661	0.0890
T=8	0.9964	0.0662	0.9031	0.0800
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.6590	0.1157	0.7401	0.8810
T=3	0.9029	0.1083	0.7535	0.3235
T=4	0.9681	0.1025	0.7520	0.1700
T=5	0.9896	0.0874	0.8014	0.1235
T=6	0.9907	0.0815	0.7930	0.1235
T=7	0.9961	0.0736	0.8176	0.1085
T=8	0.9949	0.0679	0.8404	0.0995
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.5271	0.1312	0.7623	0.9410
T=3	0.8116	0.1054	0.7216	0.6445
T=4	0.9259	0.0983	0.6902	0.3295
T=5	0.9686	0.0885	0.7031	0.2080
T=6	0.9803	0.0816	0.7110	0.1805
T=7	0.9905	0.0747	0.7308	0.1530
T=8	0.9911	0.0694	0.7510	0.1485

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 15: CRE Logit Estimates of APE_x (true value = 1), $N = 200$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9573	0.1817	0.9725	0.0635
T=3	0.9759	0.1243	0.9928	0.0555
T=4	0.9855	0.1043	0.9702	0.0650
T=5	0.9912	0.0863	1.0152	0.0480
T=6	0.9919	0.0808	0.9736	0.0600
T=7	0.9923	0.0742	0.9720	0.0630
T=8	0.9901	0.0655	1.0243	0.0510
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9594	0.1751	0.9699	0.0670
T=3	0.9739	0.1211	0.9948	0.0525
T=4	0.9851	0.1028	0.9679	0.0585
T=5	0.9899	0.0860	1.0075	0.0515
T=6	0.9906	0.0793	0.9853	0.0560
T=7	0.9927	0.0732	0.9785	0.0550
T=8	0.9905	0.0660	1.0124	0.0545
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9600	0.1701	0.9576	0.0740
T=3	0.9752	0.1168	1.0018	0.0525
T=4	0.9849	0.0998	0.9793	0.0570
T=5	0.9890	0.0856	0.9991	0.0545
T=6	0.9909	0.0782	0.9903	0.0565
T=7	0.9923	0.0720	0.9905	0.0550
T=8	0.9909	0.0657	1.0141	0.0485
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9606	0.1605	0.9715	0.0730
T=3	0.9736	0.1124	1.0085	0.0540
T=4	0.9838	0.0968	0.9878	0.0590
T=5	0.9900	0.0846	0.9965	0.0590
T=6	0.9889	0.0791	0.9733	0.0640
T=7	0.9925	0.0723	0.9845	0.0510
T=8	0.9913	0.0666	1.0011	0.0525
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9620	0.1526	0.9787	0.0705
T=3	0.9742	0.1096	1.0073	0.0600
T=4	0.9836	0.0943	1.0017	0.0530
T=5	0.9878	0.0853	0.9868	0.0635
T=6	0.9892	0.0793	0.9762	0.0660
T=7	0.9933	0.0725	0.9974	0.0550
T=8	0.9920	0.0671	1.0174	0.0560

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 16: FE OLS (LPM) Estimates of APE_x (true value = 1), $N = 200$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9572	0.1831	0.9691	0.0645
T=3	0.9765	0.1244	0.9955	0.0565
T=4	0.9854	0.1044	0.9721	0.0590
T=5	0.9911	0.0864	1.0166	0.0460
T=6	0.9919	0.0812	0.9712	0.0610
T=7	0.9920	0.0744	0.9719	0.0580
T=8	0.9900	0.0659	1.0210	0.0485
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9593	0.1764	0.9666	0.0705
T=3	0.9745	0.1210	0.9984	0.0495
T=4	0.9849	0.1029	0.9701	0.0600
T=5	0.9898	0.0859	1.0112	0.0510
T=6	0.9906	0.0797	0.9829	0.0545
T=7	0.9924	0.0734	0.9792	0.0550
T=8	0.9904	0.0663	1.0107	0.0510
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9599	0.1713	0.9546	0.0730
T=3	0.9758	0.1168	1.0057	0.0515
T=4	0.9848	0.1000	0.9807	0.0565
T=5	0.9889	0.0857	1.0012	0.0530
T=6	0.9909	0.0785	0.9886	0.0560
T=7	0.9921	0.0723	0.9898	0.0545
T=8	0.9907	0.0660	1.0124	0.0485
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9605	0.1618	0.9679	0.0715
T=3	0.9743	0.1125	1.0107	0.0530
T=4	0.9836	0.0972	0.9874	0.0610
T=5	0.9899	0.0847	0.9983	0.0585
T=6	0.9889	0.0793	0.9733	0.0635
T=7	0.9923	0.0726	0.9835	0.0535
T=8	0.9911	0.0669	0.9990	0.0540
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9617	0.1540	0.9736	0.0720
T=3	0.9748	0.1097	1.0095	0.0590
T=4	0.9834	0.0945	1.0027	0.0510
T=5	0.9877	0.0851	0.9915	0.0605
T=6	0.9892	0.0793	0.9785	0.0625
T=7	0.9930	0.0729	0.9948	0.0575
T=8	0.9918	0.0674	1.0152	0.0575

*Note: r:.05 is the rejection rate for the null of $H_0{:}\beta{=}1$ with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 17: Conditional Logit Estimates of β (true value=1), $N = 400$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	1.0263	0.1886	0.9673	0.0485
T=3	1.0109	0.1171	1.0025	0.0480
T=4	1.0039	0.0929	0.9898	0.0525
T=5	1.0076	0.0767	1.0171	0.0475
T=6	1.0018	0.0692	0.9945	0.0505
T=7	1.0018	0.0613	1.0127	0.0475
T=8	1.0005	0.0574	0.9962	0.0505
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	1.1614	0.2172	0.9537	0.0825
T=3	1.1069	0.1286	0.9936	0.0955
T=4	1.0750	0.1006	0.9699	0.0965
T=5	1.0637	0.0804	1.0166	0.1030
T=6	1.0489	0.0725	0.9865	0.0865
T=7	1.0426	0.0661	0.9713	0.1060
T=8	1.0353	0.0597	0.9859	0.0815
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	1.3578	0.2543	0.9661	0.2290
T=3	1.2540	0.1456	0.9949	0.3840
T=4	1.1908	0.1100	0.9764	0.3885
T=5	1.1589	0.0861	1.0266	0.4340
T=6	1.1301	0.0795	0.9636	0.3920
T=7	1.1122	0.0714	0.9526	0.3780
T=8	1.0968	0.0624	0.9932	0.3245
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	1.7044	0.3600	0.8957	0.5960
T=3	1.5147	0.1801	0.9895	0.8795
T=4	1.4031	0.1283	0.9908	0.9215
T=5	1.3391	0.1004	1.0139	0.9395
T=6	1.2855	0.0895	0.9703	0.9250
T=7	1.2510	0.0802	0.9515	0.9130
T=8	1.2203	0.0690	0.9977	0.9040
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	2.5486	0.6848	0.8174	0.9445
T=3	2.1272	0.2809	0.9718	1.0000
T=4	1.9192	0.1879	0.9776	1.0000
T=5	1.7912	0.1412	0.9992	1.0000
T=6	1.6870	0.1176	0.9857	1.0000
T=7	1.6150	0.1028	0.9735	1.0000
T=8	1.5523	0.0891	0.9922	1.0000

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 18: Unconditional Logit Estimates of β (true value=1), $N = 400$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	2.0525	0.3773	0.9673	0.8815
T=3	1.6217	0.2041	1.0005	0.9040
T=4	1.4147	0.1423	0.9869	0.8610
T=5	1.3152	0.1070	1.0174	0.8565
T=6	1.2445	0.0908	0.9952	0.7950
T=7	1.2024	0.0771	1.0127	0.7520
T=8	1.1711	0.0699	0.9955	0.6985
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	2.3228	0.4343	0.9537	0.9565
T=3	1.7921	0.2277	0.9907	0.9745
T=4	1.5257	0.1562	0.9678	0.9645
T=5	1.3951	0.1133	1.0180	0.9555
T=6	1.3074	0.0958	0.9881	0.9195
T=7	1.2546	0.0836	0.9710	0.8860
T=8	1.2143	0.0732	0.9844	0.8545
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	2.7155	0.5086	0.9661	0.9910
T=3	2.0584	0.2635	0.9900	0.9990
T=4	1.7102	0.1747	0.9750	0.9970
T=5	1.5329	0.1235	1.0286	0.9970
T=6	1.4172	0.1065	0.9650	0.9900
T=7	1.3444	0.0914	0.9525	0.9815
T=8	1.2909	0.0773	0.9906	0.9775
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	3.4088	0.7200	0.8957	1.0000
T=3	2.5435	0.3343	0.9872	1.0000
T=4	2.0604	0.2115	0.9907	1.0000
T=5	1.8014	0.1490	1.0174	1.0000
T=6	1.6324	0.1236	0.9694	1.0000
T=7	1.5266	0.1048	0.9523	1.0000
T=8	1.4468	0.0869	0.9952	1.0000
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	5.0971	1.3696	0.8174	1.0000
T=3	3.7240	0.5412	0.9704	1.0000
T=4	2.9656	0.3337	0.9753	1.0000
T=5	2.5184	0.2280	0.9968	1.0000
T=6	2.2179	0.1747	0.9824	1.0000
T=7	2.0248	0.1429	0.9705	1.0000
T=8	1.8804	0.1175	0.9917	1.0000

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 19: Hybrid* Conditional Logit Estimates of APE_x (true value = 1), $N = 400$

	Uncorrected				With FV Correction			
$\rho = 0.0$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5256	0.0815	1.0251	1.0000	0.7552	0.1038	0.8048	0.8050
T=3	0.6943	0.0709	1.0374	0.9895	0.9057	0.0891	0.8257	0.2910
T=4	0.7750	0.0623	1.0151	0.9480	0.9507	0.0747	0.8471	0.1800
T=5	0.8245	0.0549	1.0118	0.8855	0.9723	0.0634	0.8765	0.1160
T=6	0.8518	0.0521	0.9655	0.8275	0.9780	0.0586	0.8581	0.1170
T=7	0.8761	0.0473	0.9778	0.7545	0.9865	0.0522	0.8869	0.0925
T=8	0.8894	0.0448	0.9620	0.7145	0.9869	0.0488	0.8844	0.1015
$\rho = 0.2$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5237	0.0782	0.9164	1.0000	0.7376	0.0934	0.7677	0.9154
T=3	0.6954	0.0697	0.9734	0.9920	0.9029	0.0863	0.7862	0.3455
T=4	0.7752	0.0619	0.9689	0.9620	0.9491	0.0739	0.8111	0.1920
T=5	0.8243	0.0546	0.9799	0.8995	0.9709	0.0630	0.8496	0.1345
T=6	0.8519	0.0518	0.9412	0.8405	0.9772	0.0583	0.8373	0.1210
T=7	0.8760	0.0483	0.9341	0.7665	0.9857	0.0532	0.8481	0.1100
T=8	0.8892	0.0447	0.9474	0.7370	0.9861	0.0485	0.8721	0.1020
$\rho = 0.4$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5187	0.0733	0.7923	1.0000	0.7040	0.0798	0.7274	0.9894
T=3	0.6943	0.0679	0.8846	0.9970	0.8932	0.0821	0.7317	0.4520
T=4	0.7754	0.0608	0.9102	0.9725	0.9461	0.0721	0.7670	0.2305
T=5	0.8241	0.0535	0.9398	0.9190	0.9685	0.0616	0.8165	0.1540
T=6	0.8521	0.0521	0.8937	0.8705	0.9759	0.0585	0.7962	0.1480
T=7	0.8759	0.0489	0.8890	0.7925	0.9842	0.0539	0.8069	0.1275
T=8	0.8898	0.0440	0.9326	0.7565	0.9856	0.0479	0.8585	0.1150
$\rho = 0.6$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5079	0.0664	0.6556	1.0000	0.6296	0.0743	0.5860	1.0000
T=3	0.6923	0.0656	0.7521	0.9995	0.8710	0.0755	0.6542	0.6760
T=4	0.7751	0.0590	0.8183	0.9840	0.9385	0.0690	0.7000	0.3295
T=5	0.8231	0.0534	0.8477	0.9510	0.9633	0.0611	0.7401	0.2250
T=6	0.8515	0.0511	0.8386	0.8930	0.9722	0.0574	0.7476	0.1915
T=7	0.8763	0.0490	0.8301	0.8130	0.9824	0.0540	0.7539	0.1615
T=8	0.8903	0.0445	0.8758	0.7785	0.9842	0.0483	0.8066	0.1305
$\rho = 0.8$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.4710	0.0593	0.5687	1.0000	0.4040	0.1474	0.2286	1.0000
T=3	0.6784	0.0619	0.5772	0.9990	0.7773	0.0693	0.5156	0.9773
T=4	0.7698	0.0589	0.6315	0.9915	0.9034	0.0656	0.5667	0.6481
T=5	0.8211	0.0524	0.7012	0.9735	0.9480	0.0590	0.6228	0.4025
T=6	0.8504	0.0516	0.6971	0.9250	0.9635	0.0578	0.6224	0.3055
T=7	0.8760	0.0489	0.7215	0.8640	0.9769	0.0537	0.6562	0.2365
T=8	0.8913	0.0467	0.7382	0.8165	0.9811	0.0507	0.6796	0.2075

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors. Heterogeneity is estimated by unconditional MLE with slopes restricted to their conditional logit estimates.

Table 20: Unconditional Logit Estimates of APE_x (true value = 1), $N = 400$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.8210	0.0953	0.8365	0.6200
T=3	0.9696	0.0892	0.8782	0.1065
T=4	0.9930	0.0754	0.8902	0.0805
T=5	1.0002	0.0637	0.9093	0.0775
T=6	0.9974	0.0588	0.8818	0.0895
T=7	1.0007	0.0523	0.9066	0.0800
T=8	0.9976	0.0489	0.9001	0.0840
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.7846	0.0880	0.7995	0.8175
T=3	0.9600	0.0857	0.8519	0.1335
T=4	0.9894	0.0743	0.8643	0.0980
T=5	0.9982	0.0632	0.8887	0.0775
T=6	0.9964	0.0584	0.8652	0.0875
T=7	0.9998	0.0533	0.8702	0.0905
T=8	0.9968	0.0486	0.8895	0.0785
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.7339	0.0810	0.7612	0.9440
T=3	0.9405	0.0816	0.8092	0.2135
T=4	0.9832	0.0723	0.8337	0.1180
T=5	0.9946	0.0617	0.8665	0.0885
T=6	0.9947	0.0587	0.8302	0.0990
T=7	0.9984	0.0539	0.8344	0.1050
T=8	0.9965	0.0479	0.8799	0.0840
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.6590	0.0755	0.7549	0.9890
T=3	0.9050	0.0768	0.7469	0.4260
T=4	0.9701	0.0693	0.7894	0.1590
T=5	0.9871	0.0613	0.8071	0.1195
T=6	0.9902	0.0574	0.7954	0.1285
T=7	0.9961	0.0539	0.7917	0.1225
T=8	0.9951	0.0484	0.8348	0.1065
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.5261	0.0790	0.8153	0.9975
T=3	0.8146	0.0741	0.7019	0.8455
T=4	0.9284	0.0685	0.6950	0.3980
T=5	0.9674	0.0599	0.7322	0.2215
T=6	0.9788	0.0579	0.7051	0.1985
T=7	0.9889	0.0537	0.7198	0.1645
T=8	0.9911	0.0506	0.7290	0.1655

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 21: CRE Logit Estimates of APE_x (true value = 1), $N = 400$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9630	0.1256	0.9944	0.0660
T=3	0.9740	0.0865	1.0108	0.0595
T=4	0.9865	0.0727	0.9865	0.0600
T=5	0.9898	0.0617	1.0056	0.0480
T=6	0.9907	0.0566	0.9863	0.0535
T=7	0.9926	0.0512	0.9940	0.0590
T=8	0.9916	0.0480	0.9885	0.0590
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9613	0.1214	0.9909	0.0675
T=3	0.9741	0.0846	1.0085	0.0620
T=4	0.9862	0.0714	0.9873	0.0635
T=5	0.9890	0.0610	1.0066	0.0515
T=6	0.9906	0.0563	0.9822	0.0540
T=7	0.9923	0.0523	0.9679	0.0565
T=8	0.9913	0.0479	0.9880	0.0585
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9593	0.1140	1.0123	0.0635
T=3	0.9729	0.0824	1.0050	0.0605
T=4	0.9861	0.0690	1.0025	0.0625
T=5	0.9890	0.0598	1.0129	0.0515
T=6	0.9905	0.0568	0.9655	0.0640
T=7	0.9917	0.0529	0.9527	0.0655
T=8	0.9920	0.0472	0.9994	0.0560
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9589	0.1081	1.0218	0.0640
T=3	0.9739	0.0789	1.0182	0.0600
T=4	0.9862	0.0678	0.9995	0.0580
T=5	0.9885	0.0590	1.0124	0.0535
T=6	0.9905	0.0556	0.9791	0.0605
T=7	0.9921	0.0530	0.9501	0.0590
T=8	0.9927	0.0473	0.9997	0.0515
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9600	0.1043	1.0146	0.0720
T=3	0.9741	0.0775	1.0095	0.0590
T=4	0.9849	0.0668	1.0007	0.0570
T=5	0.9884	0.0578	1.0317	0.0515
T=6	0.9907	0.0554	0.9908	0.0550
T=7	0.9920	0.0524	0.9752	0.0570
T=8	0.9934	0.0486	0.9951	0.0565

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 22: FE OLS (LPM) Estimates of APE_x (true value = 1), $N = 400$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9627	0.1260	0.9941	0.0675
T=3	0.9738	0.0864	1.0141	0.0605
T=4	0.9862	0.0728	0.9873	0.0625
T=5	0.9896	0.0616	1.0100	0.0470
T=6	0.9905	0.0567	0.9867	0.0550
T=7	0.9925	0.0513	0.9946	0.0555
T=8	0.9914	0.0479	0.9921	0.0570
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9610	0.1218	0.9903	0.0670
T=3	0.9739	0.0846	1.0102	0.0625
T=4	0.9859	0.0715	0.9886	0.0650
T=5	0.9888	0.0608	1.0129	0.0510
T=6	0.9904	0.0564	0.9838	0.0575
T=7	0.9922	0.0525	0.9671	0.0610
T=8	0.9911	0.0479	0.9902	0.0575
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9590	0.1144	1.0120	0.0605
T=3	0.9727	0.0824	1.0074	0.0595
T=4	0.9858	0.0692	1.0022	0.0600
T=5	0.9888	0.0596	1.0181	0.0495
T=6	0.9903	0.0569	0.9657	0.0630
T=7	0.9916	0.0531	0.9518	0.0645
T=8	0.9918	0.0472	1.0014	0.0600
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9586	0.1085	1.0216	0.0675
T=3	0.9737	0.0790	1.0200	0.0570
T=4	0.9859	0.0680	0.9994	0.0570
T=5	0.9883	0.0588	1.0195	0.0495
T=6	0.9903	0.0558	0.9788	0.0600
T=7	0.9920	0.0532	0.9486	0.0605
T=8	0.9925	0.0474	1.0005	0.0545
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9597	0.1045	1.0157	0.0665
T=3	0.9739	0.0775	1.0122	0.0610
T=4	0.9846	0.0669	1.0020	0.0610
T=5	0.9882	0.0576	1.0393	0.0495
T=6	0.9905	0.0554	0.9919	0.0560
T=7	0.9919	0.0526	0.9747	0.0535
T=8	0.9932	0.0486	0.9968	0.0565

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 23: Conditional Logit Estimates of β (true value=1), $N = 600$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	1.0176	0.1516	0.9794	0.0535
T=3	1.0057	0.0956	0.9979	0.0475
T=4	0.9992	0.0744	1.0049	0.0425
T=5	1.0036	0.0642	0.9885	0.0605
T=6	1.0007	0.0555	1.0105	0.0500
T=7	0.9992	0.0507	0.9993	0.0530
T=8	1.0032	0.0463	1.0118	0.0420
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	1.1565	0.1771	0.9537	0.1165
T=3	1.1002	0.1020	1.0173	0.1270
T=4	1.0693	0.0792	1.0017	0.1145
T=5	1.0600	0.0675	0.9852	0.1375
T=6	1.0471	0.0583	0.9987	0.1170
T=7	1.0401	0.0524	0.9987	0.1060
T=8	1.0387	0.0476	1.0123	0.1200
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	1.3518	0.2121	0.9455	0.3705
T=3	1.2476	0.1164	1.0098	0.5515
T=4	1.1830	0.0873	0.9985	0.5535
T=5	1.1570	0.0741	0.9732	0.5790
T=6	1.1277	0.0624	0.9993	0.5285
T=7	1.1094	0.0560	0.9924	0.5080
T=8	1.1001	0.0505	1.0062	0.5015
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	1.6902	0.2745	0.9563	0.8195
T=3	1.5025	0.1430	1.0080	0.9710
T=4	1.3949	0.1042	0.9908	0.9875
T=5	1.3383	0.0853	0.9755	0.9890
T=6	1.2815	0.0712	0.9916	0.9850
T=7	1.2471	0.0616	1.0091	0.9845
T=8	1.2216	0.0568	0.9915	0.9830
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	2.4828	0.4973	0.8937	0.9970
T=3	2.1086	0.2235	0.9856	1.0000
T=4	1.9054	0.1540	0.9658	1.0000
T=5	1.7882	0.1174	0.9789	1.0000
T=6	1.6807	0.0948	0.9968	1.0000
T=7	1.6091	0.0804	1.0120	1.0000
T=8	1.5502	0.0723	0.9989	1.0000

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 24: Unconditional Logit Estimates of β (true value=1), $N = 600$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	2.0353	0.3032	0.9794	0.9715
T=3	1.6129	0.1669	0.9938	0.9835
T=4	1.4071	0.1135	1.0047	0.9730
T=5	1.3096	0.0896	0.9883	0.9565
T=6	1.2425	0.0729	1.0094	0.9295
T=7	1.1991	0.0638	0.9990	0.8960
T=8	1.1745	0.0564	1.0121	0.8800
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	2.3130	0.3541	0.9537	0.9955
T=3	1.7803	0.1805	1.0138	0.9985
T=4	1.5165	0.1225	1.0025	0.9975
T=5	1.3899	0.0950	0.9877	0.9915
T=6	1.3046	0.0772	0.9984	0.9860
T=7	1.2513	0.0663	0.9986	0.9780
T=8	1.2184	0.0583	1.0128	0.9720
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	2.7035	0.4241	0.9455	1.0000
T=3	2.0467	0.2102	1.0066	1.0000
T=4	1.6975	0.1385	0.9969	1.0000
T=5	1.5300	0.1062	0.9756	1.0000
T=6	1.4136	0.0837	0.9992	1.0000
T=7	1.3406	0.0717	0.9907	0.9990
T=8	1.2949	0.0624	1.0053	0.9990
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	3.3804	0.5489	0.9563	1.0000
T=3	2.5204	0.2648	1.0072	1.0000
T=4	2.0469	0.1715	0.9906	1.0000
T=5	1.8001	0.1266	0.9780	1.0000
T=6	1.6262	0.0980	0.9932	1.0000
T=7	1.5214	0.0808	1.0057	1.0000
T=8	1.4483	0.0715	0.9901	1.0000
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	4.9656	0.9945	0.8937	1.0000
T=3	3.6875	0.4301	0.9843	1.0000
T=4	2.9413	0.2730	0.9642	1.0000
T=5	2.5126	0.1893	0.9770	1.0000
T=6	2.2070	0.1403	0.9942	1.0000
T=7	2.0163	0.1113	1.0118	1.0000
T=8	1.8775	0.0956	0.9968	1.0000

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 25: Hybrid* Conditional Logit Estimates of APE_x (true value = 1), $N = 600$

	Uncorrected				With FV Correction			
$\rho = 0.0$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5214	0.0667	1.0212	1.0000	0.7512	0.0858	0.7939	0.9300
T=3	0.6919	0.0575	1.0433	0.9985	0.9027	0.0721	0.8320	0.3970
T=4	0.7724	0.0496	1.0400	0.9965	0.9478	0.0594	0.8682	0.2090
T=5	0.8226	0.0458	0.9912	0.9730	0.9702	0.0528	0.8591	0.1460
T=6	0.8521	0.0414	0.9920	0.9475	0.9784	0.0466	0.8818	0.1145
T=7	0.8732	0.0384	0.9832	0.9200	0.9832	0.0423	0.8924	0.1100
T=8	0.8919	0.0357	0.9872	0.8560	0.9896	0.0389	0.9080	0.0830
$\rho = 0.2$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5215	0.0646	0.9008	1.0000	0.7366	0.0775	0.7511	0.9750
T=3	0.6928	0.0551	1.0042	0.9995	0.8998	0.0682	0.8113	0.4695
T=4	0.7728	0.0494	0.9924	0.9955	0.9464	0.0589	0.8309	0.2440
T=5	0.8227	0.0458	0.9541	0.9735	0.9691	0.0528	0.8276	0.1700
T=6	0.8520	0.0414	0.9640	0.9540	0.9774	0.0465	0.8569	0.1345
T=7	0.8735	0.0383	0.9629	0.9230	0.9828	0.0422	0.8735	0.1145
T=8	0.8921	0.0359	0.9633	0.8615	0.9892	0.0391	0.8858	0.0890
$\rho = 0.4$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5165	0.0604	0.7799	1.0000	0.7037	0.0660	0.7140	0.9980
T=3	0.6926	0.0541	0.9058	0.9995	0.8918	0.0655	0.7491	0.5920
T=4	0.7723	0.0484	0.9332	0.9975	0.9426	0.0574	0.7869	0.3035
T=5	0.8236	0.0454	0.9059	0.9810	0.9681	0.0522	0.7878	0.1950
T=6	0.8528	0.0408	0.9320	0.9610	0.9767	0.0459	0.8293	0.1560
T=7	0.8732	0.0385	0.9234	0.9350	0.9813	0.0424	0.8383	0.1305
T=8	0.8922	0.0363	0.9260	0.8725	0.9883	0.0394	0.8513	0.1080
$\rho = 0.6$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.5058	0.0547	0.6382	1.0000	0.6328	0.0581	0.6002	1.0000
T=3	0.6896	0.0525	0.7676	1.0000	0.8690	0.0599	0.6728	0.7905
T=4	0.7724	0.0485	0.8129	0.9995	0.9356	0.0567	0.6956	0.4170
T=5	0.8235	0.0454	0.8168	0.9855	0.9641	0.0520	0.7127	0.2610
T=6	0.8514	0.0411	0.8529	0.9710	0.9724	0.0462	0.7598	0.2120
T=7	0.8730	0.0380	0.8744	0.9505	0.9788	0.0418	0.7953	0.1600
T=8	0.8917	0.0366	0.8709	0.8915	0.9858	0.0398	0.8012	0.1420
$\rho = 0.8$	Mean	SD	SE/SD	r:.05	Mean	SD	SE/SD	r:.05
T=2	0.4671	0.0484	0.5508	1.0000	0.4191	0.1050	0.2538	1.0000
T=3	0.6752	0.0504	0.5705	1.0000	0.7764	0.0545	0.5277	0.9975
T=4	0.7679	0.0491	0.6179	0.9995	0.9025	0.0547	0.5538	0.7558
T=5	0.8214	0.0448	0.6714	0.9940	0.9488	0.0505	0.5954	0.4585
T=6	0.8510	0.0418	0.7054	0.9805	0.9645	0.0469	0.6298	0.3460
T=7	0.8727	0.0388	0.7397	0.9630	0.9733	0.0427	0.6724	0.2720
T=8	0.8909	0.0374	0.7539	0.9180	0.9807	0.0405	0.6954	0.2220

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors. Heterogeneity is estimated by unconditional MLE with slopes restricted to their conditional logit estimates.

Table 26: Unconditional Logit Estimates of APE_x (true value = 1), $N = 600$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.8169	0.0786	0.8244	0.7645
T=3	0.9669	0.0720	0.8857	0.1155
T=4	0.9900	0.0600	0.9118	0.0710
T=5	0.9982	0.0532	0.8886	0.0865
T=6	0.9979	0.0469	0.9039	0.0770
T=7	0.9974	0.0424	0.9115	0.0790
T=8	1.0004	0.0389	0.9257	0.0660
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.7824	0.0728	0.7818	0.9205
T=3	0.9571	0.0676	0.8791	0.1490
T=4	0.9868	0.0593	0.8843	0.0880
T=5	0.9964	0.0530	0.8641	0.0945
T=6	0.9967	0.0467	0.8839	0.0880
T=7	0.9970	0.0423	0.8960	0.0830
T=8	1.0000	0.0391	0.9054	0.0710
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.7317	0.0674	0.7376	0.9855
T=3	0.9390	0.0646	0.8311	0.2605
T=4	0.9797	0.0576	0.8538	0.1190
T=5	0.9944	0.0524	0.8340	0.1060
T=6	0.9955	0.0461	0.8635	0.0930
T=7	0.9954	0.0425	0.8646	0.0915
T=8	0.9992	0.0394	0.8745	0.0860
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.6573	0.0634	0.7198	0.9985
T=3	0.9025	0.0607	0.7664	0.5525
T=4	0.9671	0.0568	0.7837	0.1930
T=5	0.9880	0.0522	0.7769	0.1420
T=6	0.9903	0.0462	0.8083	0.1250
T=7	0.9926	0.0419	0.8308	0.0985
T=8	0.9968	0.0397	0.8315	0.1035
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.5224	0.0660	0.7777	1.0000
T=3	0.8113	0.0593	0.7062	0.9460
T=4	0.9269	0.0566	0.6832	0.4920
T=5	0.9683	0.0512	0.7010	0.2650
T=6	0.9798	0.0471	0.7103	0.2120
T=7	0.9852	0.0427	0.7357	0.1760
T=8	0.9907	0.0405	0.7451	0.1545

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 27: CRE Logit Estimates of APE_x (true value = 1), $N = 600$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9613	0.1043	0.9788	0.0770
T=3	0.9740	0.0721	0.9902	0.0650
T=4	0.9845	0.0574	1.0213	0.0550
T=5	0.9878	0.0501	1.0102	0.0505
T=6	0.9904	0.0453	1.0048	0.0540
T=7	0.9897	0.0416	1.0004	0.0600
T=8	0.9939	0.0383	1.0129	0.0530
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9617	0.0998	0.9839	0.0805
T=3	0.9743	0.0692	1.0077	0.0685
T=4	0.9843	0.0572	1.0075	0.0610
T=5	0.9878	0.0499	1.0029	0.0485
T=6	0.9902	0.0449	1.0068	0.0565
T=7	0.9899	0.0414	1.0002	0.0620
T=8	0.9939	0.0385	1.0053	0.0505
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9613	0.0962	0.9790	0.0790
T=3	0.9732	0.0676	1.0019	0.0750
T=4	0.9842	0.0557	1.0147	0.0620
T=5	0.9883	0.0487	1.0140	0.0450
T=6	0.9910	0.0443	1.0109	0.0600
T=7	0.9896	0.0417	0.9891	0.0600
T=8	0.9939	0.0388	0.9932	0.0530
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9610	0.0925	0.9747	0.0850
T=3	0.9724	0.0658	0.9974	0.0715
T=4	0.9847	0.0552	1.0009	0.0560
T=5	0.9881	0.0484	1.0069	0.0590
T=6	0.9898	0.0445	0.9993	0.0580
T=7	0.9897	0.0414	0.9929	0.0585
T=8	0.9934	0.0393	0.9846	0.0555
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9594	0.0885	0.9761	0.0905
T=3	0.9735	0.0647	0.9880	0.0745
T=4	0.9847	0.0549	0.9944	0.0615
T=5	0.9877	0.0487	1.0004	0.0595
T=6	0.9899	0.0447	1.0025	0.0600
T=7	0.9901	0.0414	1.0083	0.0575
T=8	0.9931	0.0403	0.9811	0.0635

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.

Table 28: FE OLS (LPM) Estimates of APE_x (true value = 1), $N = 600$

$\rho = 0.0$	Mean	SD	SE/SD	r:.05
T=2	0.9610	0.1043	0.9806	0.0775
T=3	0.9738	0.0724	0.9893	0.0695
T=4	0.9841	0.0577	1.0188	0.0550
T=5	0.9876	0.0502	1.0111	0.0490
T=6	0.9903	0.0454	1.0057	0.0535
T=7	0.9896	0.0417	1.0007	0.0605
T=8	0.9937	0.0384	1.0121	0.0525
$\rho = 0.2$	Mean	SD	SE/SD	r:.05
T=2	0.9614	0.0998	0.9857	0.0780
T=3	0.9742	0.0694	1.0064	0.0675
T=4	0.9840	0.0576	1.0035	0.0575
T=5	0.9875	0.0500	1.0031	0.0500
T=6	0.9901	0.0449	1.0085	0.0540
T=7	0.9898	0.0414	1.0027	0.0625
T=8	0.9938	0.0386	1.0034	0.0540
$\rho = 0.4$	Mean	SD	SE/SD	r:.05
T=2	0.9610	0.0965	0.9795	0.0795
T=3	0.9731	0.0677	1.0016	0.0740
T=4	0.9838	0.0561	1.0098	0.0600
T=5	0.9880	0.0488	1.0151	0.0455
T=6	0.9909	0.0442	1.0153	0.0585
T=7	0.9894	0.0416	0.9920	0.0580
T=8	0.9938	0.0390	0.9920	0.0515
$\rho = 0.6$	Mean	SD	SE/SD	r:.05
T=2	0.9607	0.0927	0.9758	0.0850
T=3	0.9723	0.0660	0.9975	0.0740
T=4	0.9844	0.0556	0.9971	0.0640
T=5	0.9879	0.0486	1.0073	0.0640
T=6	0.9897	0.0445	1.0034	0.0550
T=7	0.9896	0.0414	0.9958	0.0600
T=8	0.9933	0.0394	0.9837	0.0555
$\rho = 0.8$	Mean	SD	SE/SD	r:.05
T=2	0.9592	0.0887	0.9776	0.0930
T=3	0.9733	0.0648	0.9888	0.0795
T=4	0.9843	0.0552	0.9918	0.0665
T=5	0.9875	0.0486	1.0043	0.0595
T=6	0.9898	0.0447	1.0045	0.0570
T=7	0.9900	0.0414	1.0101	0.0575
T=8	0.9929	0.0404	0.9801	0.0630

*Note: r:.05 is the rejection rate for the null of H_0:β=1 with nominal value 0.05. ρ is AR(1) serial correlation coefficient for the errors.